Patterns in Large Graphs

Tien Nam Le

Numéro National de Thèse : 2018LYSEN079

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N°512
École Doctorale en Informatique et Mathématiques de Lyon

Spécialité de doctorat : Informatique

Soutenue publiquement le 21/11/2018, par :
Tien-Nam LE

Patterns in Large Graphs

Motifs dans les grands graphes

Devant le jury composé de :

HAVET Frédéric, Directeur de Recherche, INRIA Sophia-Antipolis	Rapporteur
KRIESELL Matthias, Professeur, TU Ilmenau (Allemagne)	Rapporteur
BONAMY Marthe, Chargée de Recherche, CNRS Université Bordeaux	Examinatrice
SEBBAN Marc, Professeur, Université Jean Monnet	Examinateur
THOMASSÉ Stéphan, Professeur, ENS de Lyon	Directeur de thèse

- ÉCOLE NORMALE SUPÉRIEURE DE LYON -
Laboratoire de l'Informatique du Parallélisme (LIP) - UMR 5668

THÈSE

en vue d'obtenir le grade de

Docteur de l'Université de Lyon,
délivré par l'École Normale Supérieure de Lyon

Discipline : **Informatique**

au titre de l'École Doctorale Informatique et Mathématiques

présentée et soutenue publiquement le – Novembre 2018 par

Tien-Nam LE

PATTERNS IN LARGE GRAPHS

Directeur de thèse : Stéphan THOMASSÉ

Devant la commission d'examen formée de :

Marthe	BONAMY	*Examinatrice*
Fréderic	HAVET	*Rapporteur*
Matthias	KRIESELL	*Rapporteur*
Marc	SEBBAN	*Examinateur*
Stéphan	THOMASSÉ	*Directeur de thèse*

Contents

Acknowledgements

Printed version only...

Résumé

Un *graphe* est un ensemble de *sommets* (également appelés *noeuds*), ainsi qu'un ensemble d'*arêtes* (également appelé *liens* ou *arcs*) reliant des paires de sommets, les arêtes pouvant être orientées ou non. Les graphes ont de nombreuses applications dans les domaines des réseaux sociaux, de l'organisation des données, du traitement automatique du langage naturel, etc. Avec la quantité croissante de données collectées, il est de plus en plus important de comprendre les structures et les comportements de très grands graphes [CXWW14, JLM17, HYL17, AHKT18]. Néanmoins, l'augmentation rapide de la taille des grands graphes rend l'étude de tous les graphes de moins en moins efficace. Ainsi, il existe une demande impérieuse pour des méthodes plus efficaces pour étudier de grands graphes sans nécessiter la connaissance des graphes dans leur ensemble. Une méthode prometteuse pour comprendre les comportements de grands graphes - en particulier dans les cas liés à le flot d'informations au sein du réseau, telle que la diffusion d'informations virales ou l'adoption de nouveaux produits - consiste à exploiter les propriétés spécifiques des structures locales, telles que la taille des grappes ou la présence locale d'un *motif* spécifique - c'est-à-dire un graphe donné (généralement petit) [HJJ+18, NP12, LLDM09, Fau07]. Un exemple classique tiré de la théorie des graphes (cas avérés de la conjecture d'Erdos-Hajnal [EH89]) est que, si un graphe de grande taille ne contient pas de motif spécifique, il doit alors avoir un ensemble de sommets liés par paires ou non liés, de taille exponentiellement plus grande que prévue.

Dans cette thèse, nous aborderons certains aspects de deux questions fondamentales de la théorie des graphes concernant la présence, en abondance ou à peine, d'un motif donné H dans un grand graphe G:

(a) G peut-il être partitionné en copies de H ?

(b) Si G ne contient aucune copie de H, quelles sont les propriétés de G ?

Ces questions sont étroitement liées à certaines des conjectures les plus importantes de la théorie des graphes: les conjectures de Tutte sur les écoulements dans les graphes et la conjecture d'Erdos-Hajnal susmentionnée.

Dans la première partie, nous aborderons la question (a) - G peut-il être partitionné en copies du motif H ? - en présentant d'abord les conjectures des flots de Tutte et en montrant la connexion intime entre les flot dans les graphes et les décompositions de motifs. Nous allons prouver la conjecture de Barat-Thommasen sur les décompositions de motifs du chapitre 2 et la dépasser dans le cas où le motif est une trajectoire des chapitres 3 et 4, incluant la preuve d'une conjecture de Haggkvist et Kriesell. Nous revenons ensuite pour démontrer certaines propriétés des flot dans les graphes et un cas particulier de la conjecture de Jaeger-Linial-Payan-Tarsi au chapitre 5.

La deuxième partie est consacrée à la question (b) - Si G ne contient aucune copie du motif H, quelles sont les propriétés de G ? - où nous commençons avec la conjecture d'Erdos-Hajnal sur la non-existence de motifs dans les grands graphes et sa «dualité» dans les tournois - des graphes où chaque paire de sommets a un bord orienté. Nous montrons ensuite au chapitre 6 que la complexité locale d'un tournoi en dicte la complexité globale, ce qui résout une conjecture de Berger et al. Nous présentons un résultat majeur de Berger et al, caractérisant tous les tournois satisfaisant la conjecture d'Erdos-Hajnal dans un cas extrême, et étendons ce résultat aux graphes orientés denses du chapitre 7. Ce résultat supporter une conjecture de Harutyunian et McDiarmid que la propriété d'Erdos-Hajnal est même véritable pour les graphes dirigés en général. Enfin, au chapitre 8, nous présenterons des propriétés plus raffinées dans un cas particulier, lorsque le motif H est le triangle orienté.

Chapter 1

Introduction

1.1 Graphs and patterns

A *graph* is a set of *vertices* (also called *nodes*), together with a set of *edges* (also called *links* or *arcs*) connecting pairs of vertices, where the edges can be either oriented or non-oriented. Graphs have numerous applications in the fields of Social Networks, Data Organization, Natural Language Processing, etc. With the accumulating amount of data collected, there is a growing interest in understanding the structures and behaviors of very large graphs [CXWW14, JLM17, HYL17, AHKT18]. Nevertheless, the rapid increasing in size of large graphs makes studying the entire graphs becomes less and less efficient. Thus, there is a compelling demand for more effective methods to study large graphs without requiring the knowledge of the graphs in whole. One promising method to understand the behaviors of large graphs – especially in cases related to the flow of information among the network such as the spread of viral news or the adoption of new products – is via exploiting specific properties of local structures, such as the size of clusters or the presence locally of a specific *pattern* – i.e. a given (usually small) graph [HJJ+18, NP12, LLDM09, Fau07]. A classical example from Graph Theory (proven cases of the Erdős-Hajnal conjecture [EH89]) is that if a large graph does not contain some specific pattern, then it must have a set of vertices pairwise linked or not linked of size exponentially larger than expected.

In this thesis, we will address some aspects of two fundamental questions in Graph Theory about the presence, abundantly or scarcely, of a given pattern H in a large graph G:

(a) Can G be partitioned into copies of H?

(b) If G does not contain any copy of H, which properties does G have?

These questions are closely linked to some of the most important conjectures in Graph

Theory: the Tutte's flow conjectures on flows in graphs and the Erdős-Hajnal conjecture mentioned above.

In Part I, we will address Question (a) – Can G be partitioned into copies of pattern H? – by first presenting the Tutte's flows conjecture and showing the intimate connection between flows in graphs and pattern-decompositions. We will prove the Barát-Thommasen conjecture on pattern-decompositions in Chapter 2, and going beyond it in the case the pattern is a path in Chapters 3 and 4, including the proof of a conjecture of Häggkvist and Kriesell. We then return to prove some properties of flows in graphs and a special case of Jaeger-Linial-Payan-Tarsi's conjecture in Chapter 5.

Part II is devoted to Question (b) – If G does not contain any copy of pattern H, which properties does G have? – where we start with the Erdős-Hajnal conjecture about the non-existence of patterns in large graphs, and its "dual version" in *tournaments* – a graph where every pair of vertices has an oriented edge. We then show in Chapter 6 that the local complexity of a tournament dictates its global complexity, solving a conjecture of Berger et al. We then present a major result by Berger et al, characterizing all tournaments satisfying the Erdős-Hajnal conjecture in an extremal case, and extend that result to dense directed graphs in Chapter 7. This result supports a conjecture of Harutyunian and McDiarmid that the Erdős-Hajnal property even holds for general directed graphs. Finally, in Chapter 8, we will present some more refined properties in a specific case, when the pattern is the oriented triangle.

1.2 Some notions in Graph Theory

Graphs

A *graph* G is an ordered pair $(V(G), E(G))$ where $V(G) = \{v_1, ..., v_n\}$ is the set of *vertices* of G, and $E(G) \subseteq \{v_iv_j : 1 \leq i, j \leq n\}$ is the set of *edges* of G. The cardinality of $V(G)$ and $E(G)$ are called the *order* and *size* of G, respectively. If the edges of G are unordered, i.e. $v_iv_j = v_jv_i$ for every i, j, then G is an *undirected graph*. If the edges of G are ordered, i.e. $v_iv_j \neq v_jv_i$, then G is called a *directed graph* (or *digraph* for short). Edges in directed graphs are usually called *directed edges* or *arcs* to distinguish with edges in undirected graphs. Given an edge $e = v_iv_j$ in a graph G, v_i and v_j are called *endpoints* of e. We also say that v_i and v_j are *adjacent*, and e is *incident* with each v_i and v_j. If e is a directed edge, then v_i and v_j are the *source* and *target* of e, respectively. We also say that two edges are *incident* if they are incident to some common vertex.

A graph is *finite* if the cardinality of $V(G)$ is finite and is *infinite* otherwise. A *loop* is an edge with identical endpoints. A graph is *loopless* if it does not contain any loop. A graph G is called *simple* if it is loopless and every edge does not appear more than once in $E(G)$. In a *multigraph* G, edges and loops may appear more than once in $E(G)$. Given a directed graph G, the graph obtained from G by removing the direction of all

edges is called the *underlying graph* of G. Conversely, a directed graph obtain from an undirected graph G by adding orientations to all edges is called an *orientation* of G

Degree of vertices

Given an undirected graph G, the set of edges incident with a vertex v is denoted by $E_G(v)$, and the cardinality of $E_G(v)$ is called the *degree* of v, denoted by $d_G(v)$. A vertex with degree 0 is called an *isolated* vertex. The *minimum degree* and *maximum degree* of G are $\delta(G) := \min_{v \in V(G)} d_G(v)$ and $\Delta(G) := \max_{v \in V(G)} d_G(v)$, respectively.

Given a directed graph G, the set of edges going in (resp. going out) a vertex v is denoted by $E_G^-(v)$ (resp. $E_G^+(v)$). The *indegree* and *outdegree* of v, denoted by $d_G^-(v)$ and $d_G^+(v)$, are the cardinalities of $E_G^-(v)$ and $E_G^+(v)$, respectively. A vertex with indegre (resp. outdegree) 0 is called a *source* (resp. *sink*). We similarly have the definition of minimum/maximum in/out-degree of a vertex in a directed graph. When it is clear from the context, the subscript G can be omitted.

Subgraphs, walks and connectivity in graphs

A graph H is called a *subgraph* of another graph G, denoted by $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. In other words, H is a subgraph of G if H can be obtained from G by repeatedly removing vertices (together with incident edges) and edges. We also say that G is a *supergraph* of H or G *contains* H. If $V(H) = V(G)$, we say that H is a *spanning subgraph* of G. Given a graph G and $A \subseteq V(G)$, the subgraph obtained from G by removing all vertices in $V \setminus A$ (and their incident edges) is called the *induced subgraph* of G on A, denoted by $G[A]$.

A *walk* in a graph G is a sequence of vertices and edges $v_0, e_1, v_2, e_2, ..., e_\ell, v_\ell$ such that $e_i = v_{i-1}v_i$ for every i. The *length* of the walk is ℓ. If $v_i \neq v_j$ for every $i \neq j$, the walk is called a *path*. If $v_0 = v_\ell$ and $e_i \neq e_j$ for every $i \neq j$, the walk is called a *tour*. A tour with $v_i \neq v_j$ for every $i \neq j \leq \ell - 1$ is called a *cycle*. A tour containing every edge of G exactly once is called an *Eulerian tour*.

Two vertices u, v of an undirected graph G are *k-edge-connected* if there are k edge-disjoint paths connecting u and v in G. We say that G is *k-edge-connected* if every pair of vertices of G is k-edge-connected. If G is 1-edge-connected, we say that G is *connected* for short.

Some special graphs

The following are some special graphs; many of them will serve as patterns in the following chapters.

- A *singleton* is a graph with one vertex and no edge.

- A *path* of length ℓ, denoted by P_ℓ, is a graph with vertex set $\{v_0, ..., v_\ell\}$ and edge set $\{v_0v_1, v_1v_2, ..., v_{\ell-1}v_\ell\}$.

- A *star* with ℓ leaves, denoted by S_ℓ, is a graph with vertex set $\{v_0, v_1, ..., v_\ell\}$ and edge set $\{v_0v_1, v_0v_2, ..., v_0v_\ell\}$.

- A *cycle* (or *directed cycle* in the case the graph is directed) of length ℓ is a graph with vertex set $\{v_1, ..., v_\ell\}$ and edge set $\{v_1v_2, v_2v_3, ..., v_\ell v_1\}$.

- A *tree* is a connected undirected graph not containing any cycle. Paths and stars are trees. The vertices of a tree with degree 1 are called *leaves*.

- The *complete graph* on ℓ vertices, denoted by K_ℓ is the undirected simple graph with an edge between every pair of distinct vertices.

- A graph G is *bipartite* if $V(G)$ is the union of two disjoint sets A, B such that there is no edge between vertices of A and no edge between vertices of B. Trees are bipartite graphs.

- A *tournament* is a directed graph such that its underlying graph is a complete graph. A tournament is *transitive* if it contains no directed cycle. The transitive tournament of ℓ vertices is denoted by T_ℓ[1]. More generally, a digraph is *acyclic* if it contains no directed cycle.

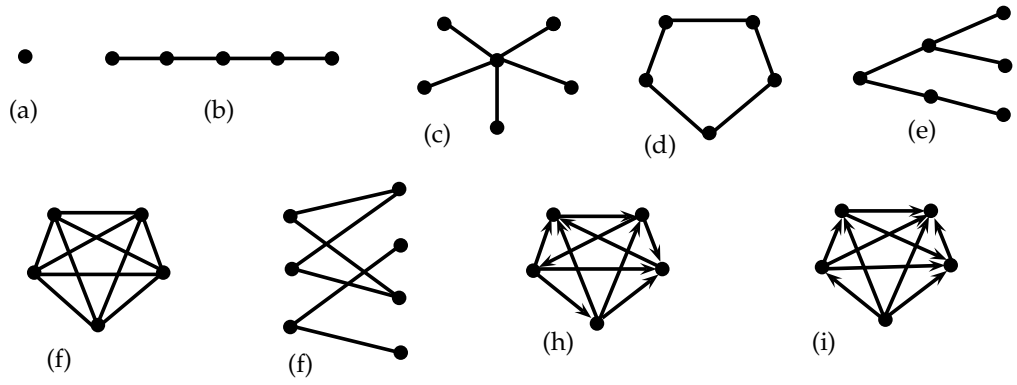

FIGURE 1.1 – List of special graphs: (a) the singleton, (b) the path of length 4 P_4, (c) the star with 5 leaves S_5, (d) the cycle of length 5 C_5, (e) a tree, (f) the complete graph of 5 vertices K_5, (g) a bipartite graph, (h) a tournament on 5 vertices, and (i) the transitive tournament on 5 vertices T_5.

[1]The usual notation in literature for the transitive tournament on ℓ vertices is TT_ℓ.

Some other notions

Two graphs G and H are *isomorphic* if there is a bijective map φ from $V(G)$ to $V(H)$ such that $uv \in E(G)$ if and only if $\varphi(u)\varphi(v) \in E(H)$. If G and H are isomorphic, we can say that H is a *copy* of G and vice versa. With the definition of isomorphic graphs, we can extend other notions, such as H is an (induced) subgraph of G if there is an (induced) subgraph of G which is a copy of H. If G does not contain H as an induced subgraph, we say that G is *H-free*.

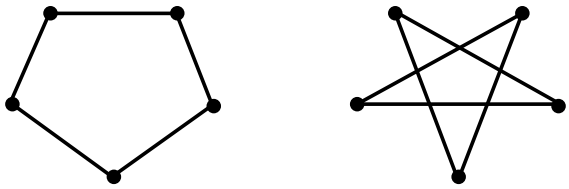

FIGURE 1.2 – The two figures in the left and right are isomorphic. Both of them are the cycle of length 5 C_5.

Given a graph G, a set of vertices $A \subseteq V(G)$ is called a *clique* if there is an edge between every pair of vertices in A, and is called an *independent set* (or *stable set*) if there is no edge between any pair of vertices in A.

A *vertex-coloring* (resp. *edge-coloring*) of a graph G is an assignment a color to each vertex (resp. edge) of G. If there are two adjacent vertices (resp. incident edges) with the same color in a vertex-coloring (resp. edge-coloring) of an undirected graph, the coloring is called *improper*. It is called *proper* otherwise.

Part I

Pattern-Decompositions of Large Graphs

Chapter 2

The Barát-Thomassen Conjecture

2.1 The Tutte's flow conjectures

In Optimization Theory or Operations Research, a *flow network* is often defined as a directed graph satisfying the *balance* property, i.e. for every vertex v (except sources and sinks) the total flows entering and leaving v are equal. Tutte's flow conjectures, however, are concerned with a more general notion of flow, where the balance property still holds but is measured on Abelian groups. Since the scope of this thesis is confined to flows on $\mathbb{Z}_p$ for some integer p, we will provide a more narrow definition of flows as follows.

DEFINITION 2.1

Given an integer p, a nowhere-zero p-flow (or p-flow for short) on a directed graph G is an assignment f of values $\{1, 2, \ldots, p-1\}$ to the edges of G such that for each vertex v,

$$\sum_{e \in E^+(v)} f(e) \equiv \sum_{e \in E^-(v)} f(e) \mod p.$$

An undirected graph G' admits a p-flow if there is an orientation G of G' admitting a p-flow.

Tutte proposed several conjectures about flows in undirected graphs, most notably the 3-flow Conjecture in 1949 and the 5-flow Conjecture in 1954 (see [Tut54], [Ste76]

The content of this chapter is covered in paper *A proof of the barát–thomassen conjecture*, Julien Bensmail, Ararat Harutyunyan, Tien-Nam Le, Martin Merker, and Stéphan Thomassé, *Journal of Combinatorial Theory, Series B*, 124:39–55, 2017.

and [BM+76]). In the rest of Part I, the term graphs is used to mention *simple, finite, and undirected graphs.*

CONJECTURE 2.2 Tutte's 3-flow Conjecture

Every 4-edge-connected graph admits a 3-flow.

CONJECTURE 2.3 Tutte's 5-flow Conjecture

Every 2-edge-connected graph admits a 5-flow.

The most well-known theorems on p-flows are the following: Seymour [Sey81] proved that that every 2-edge-connected graph has a 6-flow and Jaeger [Jae79] that every 4-edge-connected graph has 4-flow. We may easily notice that in all these conjectures and theorems, there is a strong correlation between edge-connectivity and the existence of a p-flow. One reason is that edge-connectivity requires a global connection among vertices of the graph, and so the more edge-connectivity a graph has, the easier its flow distributes among regions of the graph in order to achieve the balance property. One simple example is that a graph G made of two cliques A and B connected by a single edge e. Then G is not 2-edge-connected. No matter the assignment of value and orientation of edges, the total flow between A and B must be 0 to achieve the balance property. This implies that the value assigned to e must be 0, violating the definition of p-flow. Hence, G does not admits a p-flow.

The 3-flow conjecture is known to be notoriously hard. The weak version of it [Jae79], where 4-edge-connectivity is replaced by any arbitrarily large edge-connectivity, stood for more than 30 years.

CONJECTURE 2.4 The weak 3-flow conjecture

There exists a fixed integer k such that every k-edge-connected graph has a 3-flow.

2.2 The Barát-Thomassen conjecture

The weak 3-flow conjecture was solved by Thomassen in 2012 [Tho12] who showed that every 8-edge-connected graph has a 3-flow. His proof revealed a close link between flows in graph and pattern-decompositions. A *decomposition* of a graph G is a set of subgraphs $H_1, \ldots, H_k$ that partition the edges of G. That is, $\bigcup_{i=1}^{k} E(H_k) = E(G)$ and $E(H_i) \cap E(H_j) = \emptyset$ for every i, j. When every H_i is an isomorphic copy of a given pattern H, we call the decomposition an *H-decomposition*. Figure 2.1 shows a C_5-decomposition of the complete graph K_5.

In his proof, Thomassen [Tho12] showed that the following are equivalent:

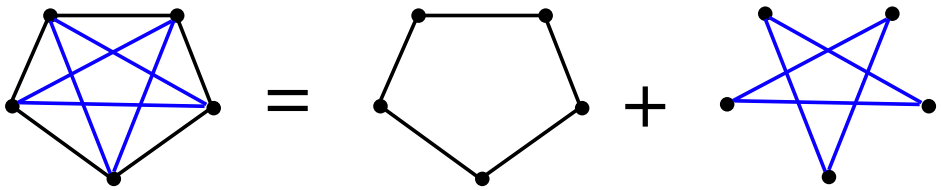

FIGURE 2.1 – A decomposition of a the complete graph K_5 into two copies of C_5.

(1) Every 8-edge-connected graph G has a 3-flow;

(2) Every 8-edge-connected graph G of size divisible by 3 has an orientation such that $d^+(v)$ is divisible by 3 for every v.

On the other hands, it is not hard to see that (2) is equivalent to the following statement,

(3) Every 8-edge-connected graph G of size divisible by 3 admits a decomposition into copies of S_3, the star of 3 leaves (note that the condition the size of G divisible by $|E(S_3)|$ is obvious and inevitable).

Indeed, suppose that (3) holds. Then for any 8-edge-connected graph G of size divisible by 3, consider an S_3-decomposition of G, and for every copy of S_3 in that decomposition, orient its edges out from the center of the star. This implies that G satisfies (2). The reverse direction is similar. Hence, (1) and (3) are equivalent, i.e., every 8-edge-connected graph has a 3-flow is equivalent to every 8-edge-connected graph has a S_3-decomposition. Said differently, Thomassen [Tho12] showed that *every highly edge-connected graph G with size divisible by $|E(S_3)|$ admits an S_3-decomposition*. This property was indeed observed several years before by Barát and Thomassen [BT06] in another paper about flows in graphs. They conjectured that the property does not only hold for S_3 but for every pattern which is a tree.

DEFINITION 2.5 The Barát-Thomassen property

> *We say that a pattern H has the* Barát-Thomassen Property *if there exists an integer k_H such that every k_H-edge-connected graph with size divisible by $|E(H)|$ has an H-decomposition.*

CONJECTURE 2.6 The Barát-Thomassen conjecture, [BT06]

> *Every tree has the Barát-Thomassen Property.*

When the conjecture was made in 2006, it was only known to hold when the pattern has less than 3 edges. The first non-trivial case was the 3-star S_3 [Tho12] that we mentioned above. In the same paper, Thomassen proved the conjecture for all stars.

Since then, the Barát-Thomassen conjecture has attracted growing attention, and it was verified for different families of trees such as *bistars*[1] [BG14, Tho13a], and paths of a certain length [BMOW16, Tho08a, Tho08b, Tho13b]. Recently, breakthrough results were obtained by Botler et al. [BMOW17], who proved the conjecture for all paths, and by Merker [Mer17], who proved the conjecture for all trees of *diameter*[2] at most 4, hence covering some of the results above.

The rest of this chapter is dedicated to provide a proof of the Barát-Thomassen conjecture.

THEOREM 2.7

Every tree has the Barát-Thomassen Property.

2.3 Proof overview

Fix a tree T with m edges as our pattern. The proof contains three steps.

(i) Given a graph G with high edge-connectivity, we first use previous results to reduce to the case where G is bipartite and one side has all degrees divisible by m. This is obtained by repeatedly removing a small number of copies of T in G.

(ii) Then we find a decomposition of G into *homomorphic copies* of T, i.e., overlapping between vertices in each copy is allowed. However we will find a way to ensure a low percentage of such bad copies.

(iii) Finally, we swap edges between homomorphic copies to convert them to isomorphic copies and complete the proof.

Simplifying G

First, it was shown by Thomassen in [Tho13a], and independently by Barát and Gerbner in [BG14], that it is sufficient to verify the Barát-Thomassen conjecture in the case G is bipartite.

THEOREM 2.8 [BG14, Tho13a]

Let T be a tree on m edges. The following two statements are equivalent:

(1) There exists a natural number k_T such that every k_T-edge-connected graph with size divisible by m has a T-decomposition.

[1] A *bistar* is a tree with diameter at most 3.
[2] The *diameter* of a graph G is the maximum length of a path in G.

(2) *There exists a natural number k'_T such that every k'_T-edge-connected bipartite graph with size divisible by m has a T-decomposition.*

A corollary of the Barát-Thomassen conjecture for stars [Tho12] is the following decomposition theorem, which was shown by Thomassen in [Tho13a] and also applied in [BMOW16] and [Mer17].

THEOREM 2.9 [Tho13a]

Let G be a bipartite graph with partition classes A_1 and A_2, and size divisible by m. If G is $(4\lambda + 6m)$-edge-connected, then G can be decomposed into two λ-edge-connected graphs G_1 and G_2 such that $d_{G_i}(v)$ is divisible by m for every v in A_i.

By Theorems 2.8 and 2.9, it is sufficient to prove the Barát-Thomassen conjecture for bipartite graphs G on vertex classes A and B, where all vertices in A have degree divisible by m, the size of the pattern T.

Equitable coloring

Since every tree is bipartite, let us call T_A and T_B the vertex classes of the bipartition of T, and we may assume that T_B contains a leaf. The T-decompositions we are going to construct will respect the bipartitions of G and T in the sense that the vertices corresponding to T_A will lie in A for each copy of T. We say that vertices $v \in V(G)$ and $t \in V(T)$ are *compatible* if $v \in A$ and $t \in T_A$, or $v \in B$ and $t \in T_B$.

We consider a specific kind of edge-coloring of G, that was introduced in [Mer17]. Assuming G is (improperly) edge-colored, we denote by $d_i(v)$ the degree of vertex v in color i. For $t \in V(T)$, let $E_T(t)$ denote the set of edges incident with t.

DEFINITION 2.10

An edge-coloring $\phi : E(G) \to E(T)$ is called T-equitable, if for any compatible vertices $v \in V(G), t \in V(T)$ and $j, k \in E_T(t)$, we have $d_j(v) = d_k(v)$.

It was shown by Merker [Mer17] that highly edge-connected graphs admit T-equitable edge-colorings.

THEOREM 2.11 Theorem 3.4 in [Mer17]

For all natural numbers m and L there exists a natural number $f(m, L)$ such that the following holds: If G is an $f(m, L)$-edge-connected bipartite graph with bipartite classes A and B where all vertices in A have degree divisible by m, and T

> *is a tree on m edges with bipartite classes T_A and T_B where T_B contains a leaf, then G admits a T-equitable coloring where the minimum degree in each color is at least L.*

Notice that since we put no constraints on the degrees in B, it is necessary that the greatest common divisor of the degrees in T_B is 1 if we want to construct a T-equitable coloring. For this reason we chose the bipartition of T so that T_B contains a leaf.

If there exists a T-decomposition of a bipartite graph G where all copies of T are oriented the same way (with respect to the bipartite classes), then this gives rise to a T-equitable coloring of G. Vice versa, a T-equitable coloring can also be used to construct a T-decomposition. This was done in [Mer17] for the case that the girth of G is at least the diameter of T, and also in general for trees of diameter at most 4.

We then use probabilistic methods to show that a T-equitable coloring can be turned into a T-decomposition whenever the minimum degree in each color is large enough.

THEOREM 2.12

> *Let T be a tree on m edges and let G be a bipartite graph admitting a T-equitable coloring. If the minimum degree in each color is at least 10^{50m}, then G has a T-decomposition.*

Combined with the previous theorems, Theorem 2.12 completes the proof of the Barát-Thomassen conjecture as follows.

Proof of Theorem 2.7: By Theorem 2.8, we may assume that G is bipartite. We show that every $(4f(m, 10^{50m}) + 6m)$-edge-connected bipartite graph has a T-decomposition, where f is the function given by Theorem 2.11. By Theorem 2.9 we can decompose G into two spanning $f(m, 10^{50m})$-edge-connected graphs G_1 and G_2, such that in one side of the bipartition of each G_i all vertices have degree divisible by m. By Theorem 2.11, we can find a T-equitable coloring of G_i in which the minimum degree in each color is at least 10^{50m}. This coloring can be turned into a T-decomposition by Theorem 2.12. □

Pseudo-decompositions and the proof of Theorem 2.12

In our proof of Theorem 2.12, a T-decomposition of a graph G is obtained in two steps, which we describe more formally below.

- In the first step we construct a decomposition of G into so-called *pseudo-copies* of T, which are subgraphs that are in some sense homomorphic to T. Such a decomposition, which we call a *T-pseudo-decomposition*, can easily be obtained

from a T-equitable coloring, see also [Mer17]. Instead of choosing any such decomposition, we use probabilistic methods to find one in which the vast majority of pseudo-copies at every vertex are isomorphic to T. The details of this step can be found in Section 2.4.

- In the second step, we use these isomorphic copies to repair the non-isomorphic copies of T by making subgraph switches. While the switching itself is a deterministic operation, we again use probabilistic methods to find a suitable set of isomorphic copies. This part of the proof is detailed in Section 2.5.

Step 1: Finding a good T-pseudo-decomposition

Let G be a graph with a T-equitable coloring. Recall that $E_T(t)$ denotes the set of edges incident with a vertex $t \in V(T)$. Let us denote by $N_i(v)$ the set of edges colored i incident with $v \in V(G)$. Furthermore, we set $N_{E_T(t)}(v) := \bigcup_{i \in E_T(t)} N_i(v)$ for every $t \in V(T)$ compatible with v. Since the edge-coloring of G is T-equitable, we have that $|N_{E_T(t)}(v)|$ is divisible by $|E_T(t)|$ for every compatible $v \in V(G)$, $t \in V(T)$. Thus, we partition $N_{E_T(t)}(v)$ into stars of size $|E_T(t)|$ that contain each of the colors in $E_T(t)$ exactly once. Let $\mathcal{S}$ be the collection of stars we get after having done this for every $v \in V(G)$ and compatible $t \in V(T)$. Consider an auxiliary graph $G_{\mathcal{S}}$ whose vertices are the stars in $\mathcal{S}$, and where two vertices are joined by an edge whenever the corresponding stars have an edge in common. By construction, each connected component of $G_{\mathcal{S}}$ is a tree isomorphic to T. For every connected component in $G_{\mathcal{S}}$, we take the union of all the stars corresponding to it in G. It is easy to see that this decomposes G into parts of the same size as T. In fact even more is true: Each part is isomorphic to a graph obtained from T by identifying vertices. This motivates the following definition.

DEFINITION 2.13

A graph H is a pseudo-copy of T, if there exists a surjective graph homomorphism $h : V(T) \to V(H)$ that induces a bijection between $E(T)$ and $E(H)$.

In other words, a graph H is a pseudo-copy of T, if it is isomorphic to a multigraph obtained from T by identifying vertices and keeping all edges. We also refer to pseudo-copies of T as *pseudo-trees*. The term *T-pseudo-decomposition* denotes a decomposition where each part is a pseudo-copy of T. Given a T-equitable coloring of G, the construction above results in a T-pseudo-decomposition of G.

Notice that it might be the case that a graph H can be considered as a pseudo-copy of T in different ways if there exists more than one homomorphism from T to H with the required properties. However, we will only consider homomorphisms that induce the same edge-coloring of H as the given T-equitable coloring. Furthermore, we only consider pseudo-copies of T in G that respect the bipartition in the sense that vertices corresponding to T_A always lie in A.

Let $\mathcal{P}$ be a T-pseudo-decomposition of G. For every compatible $v \in V(G)$ and $t \in V(T)$, we denote by $N_{\mathcal{P}}(v|t)$ the set of pseudo-trees in $\mathcal{P}$ in which v is the image of t. Let $d_{\mathcal{P}}(v|t) = |N_{\mathcal{P}}(v|t)|$. Clearly, for any two different vertices u and v of G, we have $N_{\mathcal{P}}(u|t) \cap N_{\mathcal{P}}(v|t) = \emptyset$. Notice also that

$$\bigcup_{v \in G} N_{\mathcal{P}}(v|t) = \mathcal{P}$$

for every $t \in V(T)$.

So far we have explained how T-pseudo-decompositions can be obtained from a T-equitable coloring. We denote such a resulting T-pseudo-decomposition $\mathcal{P}$ of G by $\mathcal{H} \cup \mathcal{I}$, where $\mathcal{I}$ denotes the collection of pseudo-copies that are isomorphic to T and $\mathcal{H}$ denotes the collection of the remaining pseudo-copies.

If the minimum degree in each color is large in the T-equitable coloring, then there are many possibilities at every vertex to decompose its incident edges into stars. Using probabilistic methods, we find a T-pseudo-decomposition where $d_{\mathcal{H}}(v|t) \leqslant \varepsilon d_{\mathcal{I}}(v|t)$ for some given $\varepsilon > 0$ and every compatible $v \in V(G)$, $t \in V(T)$. Now for every non-isomorphic copy $H \in N_{\mathcal{H}}(v|t)$, there are many copies isomorphic to T in $N_{\mathcal{I}}(v|t)$. We will use one of these isomorphic copies to improve the T-pseudo-decomposition by repairing H. This is done by a *subgraph switch operation* which is explained in more detail in Step 2. However, if the trees in $N_{\mathcal{I}}(v|t)$ overlap too much, then we might not be able to make any switch that improves the T-pseudo-decomposition. To avoid this, we need to find a large set of isomorphic copies in $N_{\mathcal{I}}(v|t)$ that pairwise intersect only in v. To measure how much the pseudo-trees in a T-pseudo-decomposition overlap, we use the following concept called conflict ratio.

DEFINITION 2.14

Let $\mathcal{P}$ be a collection of pseudo-copies of T in G, and $v \in V(G)$ and $t \in V(T)$ be compatible vertices. The conflict ratio *of v with respect to t, denoted by* $\mathrm{conf}_{\mathcal{P}}(v|t)$, *is defined by*

$$\mathrm{conf}_{\mathcal{P}}(v|t) := \frac{\max_{u \neq v} \left|\{T \in N_{\mathcal{P}}(v|t) : u \in V(T)\}\right|}{d_{\mathcal{P}}(v|t)}.$$

Intuitively, $\mathrm{conf}_{\mathcal{P}}(v|t)$ measures the maximum proportion of pseudo-copies in $N_{\mathcal{P}}(v|t)$ in which some fixed vertex u appears. Clearly, we always have $0 \leqslant \mathrm{conf}_{\mathcal{P}}(v|t) \leqslant 1$. If v and t are not compatible, then we set $\mathrm{conf}_{\mathcal{P}}(v|t) = 0$. Globally, we define $\mathrm{conf}(\mathcal{P}|t) := \max_v \mathrm{conf}_{\mathcal{P}}(v|t)$ and $\mathrm{conf}(\mathcal{P}) := \max_t \mathrm{conf}(\mathcal{P}|t)$.

To ensure that the isomorphic copies in the T-pseudo-decomposition $\mathcal{H} \cup \mathcal{I}$ are sufficiently spread out, we also require $\mathrm{conf}(\mathcal{I}) \leqslant \delta$ for some given $\delta > 0$. In Section 2.4, we prove that such a T-pseudo-decomposition can always be obtained provided the minimum degree in each color is large enough.

LEMMA 2.15

Let T be a tree on m edges and ε, δ real numbers with $0 < \varepsilon, \delta < 1$. Let G be a T-equitably colored bipartite graph where the minimum degree in each color is at least $(10m)^{18}(\varepsilon\delta)^{-6}$. Then G admits a T-pseudo-decomposition $\mathcal{H} \cup \mathcal{I}$, where $\mathcal{I}$ denotes the collection of isomorphic copies of T, such that:

(1) for every compatible $v \in V(G)$ and $t \in V(T)$, we have $d_{\mathcal{H}}(v|t) \leqslant \varepsilon d_{\mathcal{I}}(v|t)$;

(2) $\mathrm{conf}(\mathcal{I}) \leqslant \delta$.

Step 2: Repairing non-isomorphic copies

For this part of the proof we label the vertices $t_0, \dots, t_m$ of T so that, for every $i \in \{1, \dots m\}$, the subgraph induced by $t_0, \dots, t_i$ is connected. Such an ordering can for example be obtained by applying a breadth-first search algorithm from some vertex t_0 of T. We also label the edges of T so that e_i denotes the edge joining t_i with $T[t_0, \dots, t_{i-1}]$ for every $i \in \{1, \dots m\}$. To indicate at which place a pseudo-copy H fails to be isomorphic to T, we introduce the following definitions.

DEFINITION 2.16

Let H be a pseudo-copy of T, and let v_i denote the image of t_i in H for every $i \in \{0, \dots, m\}$. For $i \in \{1, \dots, m\}$, we say that H is i-good if the vertices $v_0, \dots, v_i$ are pairwise distinct. If H is not i-good, then we say that H is i-bad.

Note that since G does not have multiple edges, every pseudo-copy of T in G is 2-good. Moreover, since G is bipartite, every pseudo-copy of T in G is even 3-good.

The idea is to use isomorphic copies to repair the pseudo-trees that are not isomorphic to T. We start by considering all pseudo-trees in $\mathcal{H}$ that are 4-bad. For each such H, we will find an isomorphic copy $f(H)$ in $\mathcal{I}$ such that $H \cup f(H)$ can be written as the union of two 5-good pseudo-copies of T, say $H_1 \cup H_2$. We then remove H from $\mathcal{H}$ and $f(H)$ from $\mathcal{I}$, and add $\{H_1, H_2\}$ to $\mathcal{H}$. The technical definition of this so-called *switch* is given below. We use this operation for all 4-bad pseudo-copies of T in $\mathcal{H}$. Let $\mathcal{H}' \cup \mathcal{I}'$ denote the resulting T-pseudo-decomposition, where $\mathcal{I}'$ again contains only isomorphic copies of T and all pseudo-copies in $\mathcal{H}'$ are 4-good. We repeat this step, this time repairing all 5-bad pseudo-copies in $\mathcal{H}'$ by using isomorphic copies in $\mathcal{I}'$. We continue like this until we get a T-pseudo-decomposition in which all pseudo-copies are m-good and thus isomorphic to T.

To make sure that we can perform a switch between H and $f(H)$, we need $f(H)$ to satisfy certain properties. Let v_j denote the image of t_j in H for $j \in \{0, \dots, m\}$ and suppose i is chosen minimal such that H is i-bad. By the choice of our labelling, there exists $i' \in \{0, \dots, i-1\}$ with $t_{i'}t_i \in E(T)$. To ensure that v_i is distinct from the

previous vertices $v_0, \dots, v_{i-1}$, we want to choose a different edge corresponding to e_i at $v_{i'}$. Since we take this edge from $f(H)$, we want $f(H)$ to also use the vertex $v_{i'}$ as image of t'_i. However, this should be the only point of intersection with H to ensure that both copies will be i-good after the switch.

More precisely, for every edge $e_i \in E(T)$, let T^{i-} denote the connected component of $T - e_i$ containing t_0. Let T^{i+} be the subgraph of T induced by $E(T) \setminus E(T^{i-})$. If H is a pseudo-copy of T, then we denote the images of T^{i-} and T^{i+} under the homomorphism by H^{i-} and H^{i+}. Now we are ready to define the switching operation.

DEFINITION 2.17

Let $\mathcal{H}$ be a collection of pseudo-copies of T in G and $i \in \{1, \dots, m\}$. Let $t_{i'}$ be the endpoint of the edge e_i that is different from t_i. Suppose $H_1, H_2 \in N_{\mathcal{H}}(v|t_{i'})$ for some $v \in V(G)$. The i-switch of $\{H_1, H_2\}$ is defined by

$$\mathrm{sw}_i(\{H_1, H_2\}) = \{H_1^{i+} \cup H_2^{i-}, H_1^{i-} \cup H_2^{i+}\}.$$

By making an i-switch between two pseudo-copies H and $f(H)$, their vertices corresponding to $v_0, \dots, v_{i-1}$ remain unchanged. In particular, if both H and $f(H)$ are $(i-1)$-good, then also both copies in $\mathrm{sw}_i(\{H, f(H)\})$ will be $(i-1)$-good. Moreover, if $H \cap f(H) = \{v_{i'}\}$, then after the switch both pseudo-trees will be i-good. Notice that neither of the two new pseudo-trees is necessarily still isomorphic to T. In particular, the collection of isomorphic copies might shrink with every step of the repairing process.

If the pseudo-trees in $\mathcal{I}$ overlap too much, we might not be able to find a single pseudo-tree $f(H)$ in $\mathcal{I}$ with $H \cap f(H) = \{v_{i'}\}$. A sufficiently low conflict ratio of $\mathcal{I}$ ensures that we can find such a function $f : \mathcal{H} \to \mathcal{I}$. However, to continue this process we also need that the remaining collection of isomorphic copies $\mathcal{I} \setminus f(\mathcal{H})$ has a low conflict ratio. To this end we use the Local Lemma to prove the following lemma in Section 2.5.

LEMMA 2.18

Let T be a tree on m edges and ε, δ positive real numbers with $\varepsilon + \delta m < \frac{1}{2}$ and $\varepsilon m < 1$. Let $\mathcal{H}$ and $\mathcal{H}'$ be collections of pseudo-copies of T in G with $\mathrm{conf}(\mathcal{H}') \leqslant \delta$ and $d_{\mathcal{H}'}(v|t) > \max\{22/\varepsilon^7, d_{\mathcal{H}}(v|t)/\varepsilon\}$ for each compatible $v \in V(G)$, $t \in V(T)$. For every $t \in V(T)$, there exists an injective function $f_t : \mathcal{H} \to \mathcal{H}'$ such that

- *$f_t(N_{\mathcal{H}}(v|t)) \subset N_{\mathcal{H}'}(v|t)$ for every $v \in V(G)$ compatible with t,*
- *$H \cap f_t(H) = \{v\}$ for every $H \in N_{\mathcal{H}}(v|t)$, and*
- *$d_{f_t(\mathcal{H})}(v|t') \leqslant 3\varepsilon d_{\mathcal{H}'}(v|t')$ for every compatible $v \in V(G)$, $t' \in V(T)$.*

By using Lemma 2.18 with $\mathcal{H}' = \mathcal{I}$, we find a collection $f(\mathcal{H})$ in which the degrees are low compared to the degrees in $\mathcal{I}$. Thus, the conflict ratio of the collection of isomorphic copies only increases by a constant factor after each step of the repairing process. By choosing ε and δ sufficiently small, the proof of Theorem 2.12 will follow from Lemma 2.15 and repeated applications of Lemma 2.18. The details can be found at the end of Section 2.5.

2.4 Find a good pseudo-decomposition

Given a graph with a T-equitable coloring and large minimum degree in each color, we construct a T-pseudo-decomposition satisfying the conditions in Lemma 2.15. As described in Step 1 of Section 2.3, every T-equitable coloring gives rise to several T-pseudo-decompositions. We form the pseudo-copies of T by grouping the edges at every vertex randomly into rainbow stars. If the degrees in each color are large enough, we can ensure that most of the resulting pseudo-trees are isomorphic to T and also the conflict ratio of the resulting T-pseudo-decomposition is small. The proof of this is essentially an application of the Local Lemma.

PROPOSITION 2.19 Symmetric Local Lemma

> *Let $A_1, ..., A_n$ be events in some probability space Ω with $\mathbb{P}[A_i] \leqslant p$ for all $i \in \{1, \dots, n\}$. Suppose that each A_i is mutually independent of all but at most d other events A_j. If $4pd < 1$, then $\mathbb{P}[\cap_{i=1}^{n} \overline{A_i}] > 0$.*

The bad events in this case are of the form that many copies in $N_{\mathcal{H}}(v|t)$ are either not isomorphic to T or have a vertex different from v in common. To show that each event occurs with low probability, we make use of an inequality due to McDiarmid [McD02] (see also [MR13]).

PROPOSITION 2.20 McDiarmid's Inequality (simplified version)

> *Let X be a non-negative random variable, not identically 0, which is determined by m independent permutations $\Pi_1, ..., \Pi_m$. A* choice *is the position that a particular element gets mapped to in a permutation. If there exist $d, r > 0$ such that*
>
> - *interchanging two elements in any one permutation can affect X by at most d, and*
> - *for any $s > 0$, if $X \geqslant s$ then there is a set of at most rs choices whose outcomes certify that $X \geqslant s$,*

then for any $0 \leqslant \lambda \leqslant \mathbb{E}[X]$,

$$\mathbb{P}\left[|X - \mathbb{E}[X]| > \lambda + 60d\sqrt{r\mathbb{E}[X]}\right] \leqslant 4e^{-\frac{\lambda^2}{8d^2r\mathbb{E}[X]}}.$$

A necessary condition to apply the Local Lemma is that each event is mutually independent of most other events. To make sure that this is the case, we start the proof by partitioning the edges at each vertex into so-called *fans* of roughly the same size. Recall that for $v \in V(G)$ and $t \in V(T)$, we denote by $N_i(v)$ the edges colored i incident with v in G, and by $E_T(t)$ the set of edges incident with t in T.

Proof of Lemma 2.15: Set $c = \lceil (10m)^9(\varepsilon\delta)^{-3} \rceil$. For every $v \in V(G)$ and color i, we choose $r_{v,i} \in \{0, \dots, c-2\}$ such that $d_i(v) \equiv r_{v,i} \pmod{c-1}$. Since the minimum degree in each color in G is greater than c^2, we can partition every set $N_i(v)$ into subsets of size c and $c-1$ so that precisely $r_{v,i}$ of them have size c. We call these subsets *i-blades*. Note that an edge uv of color i in G appears both in an i-blade of $N_i(u)$ as well as in an i-blade of $N_i(v)$, but we do not require these two i-blades to have the same size.

For every compatible $t \in V(T)$, $v \in V(G)$, and $i, j \in E_T(t)$, we have $d_i(v) = d_j(v)$ since the coloring is T-equitable. Thus, the number of i-blades of size c (respectively, of size $c-1$) in the partition of $N_i(v)$ is equal to the number of j-blades of size c (respectively, of size $c-1$) in the partition of $N_j(v)$. We can therefore partition the edges of $N_{E_T(t)}(v)$ into *fans*, which are unions of blades of the same size, such that every fan contains precisely one i-blade for every $i \in E_T(t)$. In other words, a fan φ at a vertex v (with relation to t) is a subset of $N_{E_T(t)}(v)$ of size $c|E_T(t)|$ or $(c-1)|E_T(t)|$ such that all colors in $E_T(t)$ appear c times or $c-1$ times in φ. We also call φ a *t-fan* to indicate the colors appearing in φ.

For every compatible $t \in V(T)$, $v \in V(G)$, and every t-fan φ at v, we uniformly at random partition the edges in φ into rainbow stars of size $|E_T(t)|$. More precisely, for every $i \in E_T(t)$ we choose a permutation $\Pi_{\varphi,i}$ independently and uniformly at random from all permutations on c elements (respectively, on $c-1$ elements if the blades of φ have size $c-1$). By labelling the edges of each blade, each permutation $\Pi_{\varphi,i}$ corresponds to an ordering of the edges of the i-blade of φ. Now we partition the edges of φ into stars of size $|E_T(t)|$ by grouping the edges of different blades that were mapped to the same position. In other words, for every $s \in \{1, \dots, c\}$ (respectively, $s \in \{1, \dots, c-1\}$) we form a star by choosing for every $i \in E_T(t)$ the edge labeled $\Pi_{\varphi,i}(s)$ in the i-blade of φ. These stars are centered at v and each color in $E_T(t)$ appears precisely once. Note that every edge $uv \in E(G)$ belongs to exactly two stars, one centered at u and one centered at v. As described in Step 1 in Section 2.3, these stars correspond to a T-pseudo-decomposition of G in a canonical way. All that remains to show is that there exists an outcome of the random permutations such that the resulting T-pseudo-decomposition satisfy the conditions (1) and (2) of Lemma 2.15.

We denote the set of pseudo-trees using edges of a fan φ by $\mathcal{T}_\varphi$. Note that $|\mathcal{T}_\varphi|$ is either equal to $c-1$ or c. Now we formally define what the bad events at a t-fan φ at a vertex v are. Let A_φ be the event that more than $2m^2c^{2/3}$ of the pseudo-copies in $\mathcal{T}_\varphi$ are not isomorphic to T. Let B_φ be the event that there exists a vertex $u \in V(G)$ with $u \neq v$ such that more than $2mc^{2/3}$ pseudo-copies in $\mathcal{T}_\varphi$ contain u. Finally, let $C_\varphi = A_\varphi \cup B_\varphi$. We will prove the following two statements.

Claim 1: Each C_φ is mutually independent of all but at most $4(cm)^{2m}$ other events C_ψ.

Claim 2: $\mathbb{P}[C_\varphi] < 9(cm)^m m e^{-c^{2/3}/32}$.

Before we proceed to prove these claims, let us note that they allow us to use the Local Lemma to get our desired T-pseudo-decomposition $\mathcal{H} \cup \mathcal{I}$. Indeed, since $e^{-x} < \frac{(9m)!}{x^{9m}}$ for $x > 0$, we have

$$\begin{aligned}
4 \cdot 4(cm)^{2m} \cdot \mathbb{P}[C_\varphi] &< 2^8 \cdot (cm)^{3m} \cdot m \cdot e^{-c^{2/3}/32} \\
&< 2^{45m+8} \cdot \left(\frac{m}{c}\right)^{3m} \cdot m \cdot (9m)! \\
&< \left(2^{18} \cdot \frac{m}{c} \cdot (9m)^3\right)^{3m} \\
&< \left(\frac{10^9 m^4}{c}\right)^{3m} \\
&< 1\,,
\end{aligned}$$

where the last inequality follows from $c \geqslant (10m)^9$. Thus, the symmetric version of the Local Lemma yields a T-pseudo-decomposition $\mathcal{H} \cup \mathcal{I}$ for which none of the events C_φ holds. Now $\mathcal{H} \cup \mathcal{I}$ has the desired properties:

- Since A_φ does not hold for any φ, at most $2m^2c^{2/3}$ of the pseudo-copies in $\mathcal{T}_\varphi$ are not isomorphic to T. Since $c \geqslant (10m)^9\varepsilon^{-3}$, we have $2m^2c^{2/3} < \frac{\varepsilon}{1+\varepsilon}c$. Thus, less than $\frac{\varepsilon}{1+\varepsilon}c$ of the pseudo-copies in $\mathcal{T}_\varphi$ are in $\mathcal{H}$, while at least $\frac{1}{1+\varepsilon}c$ of them are in $\mathcal{I}$. This holds for every t-fan at v, so we have $d_{\mathcal{H}}(v|t) < \varepsilon d_{\mathcal{I}}(v|t)$.

- Since B_φ does not hold for any φ, there are at most $2mc^{2/3}$ trees in $\mathcal{T}_\varphi$ containing a given vertex u different from v. As argued above, at least $\frac{c}{1+\varepsilon}$ of the pseudo-copies in $\mathcal{T}_\varphi$ are in $\mathcal{I}$. Since $c \geqslant (10m)^9(\varepsilon\delta)^{-3}$, we have $2mc^{2/3} < \delta\frac{c}{1+\varepsilon}$. Thus, the proportion of trees in $\mathcal{T}_\varphi \cap \mathcal{I}$ containing u is less than δ. This is true for every t-fan at v, so we have

$$\frac{\big|\{H \in N_{\mathcal{I}}(v|t) : u \in V(H)\}\big|}{d_{\mathcal{I}}(v|t)} \leqslant \delta$$

  and thus $\mathrm{conf}(\mathcal{I}) \leqslant \delta$.

It remains to verify Claims 1 and 2. We begin by proving Claim 1.

Proof of Claim 1: The structure of $\mathcal{T}_\varphi$ depends on permutations in different fans. Let $J(\varphi)$ denote the set of fans ψ for which there exists an outcome of the random permutations such that $\mathcal{T}_\varphi \cap \mathcal{T}_\psi$ is non-empty. Since each fan consists of at most cm edges, there are at most $cm + (cm)^2 + \ldots + (cm)^m$ fans we can reach from φ via a path of length at most m. Thus,

$$|J(\varphi)| \leqslant cm + (cm)^2 + \ldots + (cm)^m < 2(cm)^m .$$

This shows that there are at most $2(cm)^m$ fans where the outcome of the permutation affects the structure of $\mathcal{T}_\varphi$. The same calculation shows that each permutation affects the structure of at most $2(cm)^m$ sets $\mathcal{T}_\psi$. Hence, the event C_φ is mutually independent of all but at most $4(cm)^{2m}$ other events C_ψ. □

Before we prove Claim 2, let us introduce more terminology. Let t_i and t_j be two distinct vertices of T. Notice that t_i or t_j could be equal to t. We say that a pseudo-copy H of T is *(t_i, t_j)-bad* if the images of t_i and t_j in H are identical. For a t-fan φ at a vertex v, let $A_\varphi(t_i, t_j)$ be the event that the number of (t_i, t_j)-bad pseudo-trees in $\mathcal{T}_\varphi$ is greater than $2c^{2/3}$. For a vertex $u \in V(G)$ with $u \neq v$, let $B_\varphi(u|t_i)$ be the event that the number of pseudo-trees in $\mathcal{T}_\varphi$ in which u is the image of t_i is greater than $2c^{2/3}$. The proof of Claim 2 consists of two parts:

Claim 2A: $\mathbb{P}[A_\varphi(t_i, t_j)] < 4e^{-c^{2/3}/32}$ for every $t_i, t_j \in V(T)$ with $t_i \neq t_j$.

Claim 2B: $\mathbb{P}[B_\varphi(u|t_i)] < 4e^{-c^{2/3}/8}$ for every $u \in V(G)$, $t_i \in V(T)$ and $u \neq v$.

The proofs of Claims 2A and 2B use McDiarmid's inequality and have a very similar structure. We will therefore present all the details in the proof of Claim 2A, and only point out the differences in the proof of Claim 2B.

Proof of Claim 2A:. Fix t_i and t_j as different vertices of T. Let P_i and P_j denote the paths in T from t to t_i and t_j. In the case that one is a subpath of the other, we may assume that P_i is contained in P_j. Let $t_{j'}$ denote the second last vertex of P_j and let j denote the edge joining $t_{j'}$ and t_j. Now $T - j$ consists of two components, one of which contains t_j while the other one contains t, t_i, and $t_{j'}$.

Let π be a fixed outcome of all permutations apart from those at the j-blades of $t_{j'}$-fans. In other words, given π, we only need to know the outcome of the permutations $\Pi_{\psi,j}$ for every $t_{j'}$-fan ψ to construct the T-pseudo-decomposition. For any such outcome π, we will show that the conditional probability $\mathbb{P}[A_\varphi(t_i, t_j)|\pi]$ is at most $4e^{-c^{2/3}/32}$. Clearly, since we condition on an arbitrary but fixed event, this uniform bound implies Claim 2A.

Let T' denote the component of $T - j$ containing t, t_i and $t_{j'}$, and let T'' denote the subgraph of T induced by $E(T) \setminus E(T')$. Let $\mathcal{T}'_\varphi$ denote the images of T' in the pseudo-trees of $\mathcal{T}_\varphi$. By fixing π, the set $\mathcal{T}'_\varphi$ is also fixed. The permutations of the j-blades at the $t_{j'}$-fans only decide how the images of T' and T'' get matched at the $t_{j'}$-fans.

Let Ψ denote the set of $t_{j'}$-fans which contain edges of pseudo-copies in $\mathcal{T}_\varphi$. Note that also the set Ψ is completely determined by π. Let X_φ denote the random variable counting the number of (t_i, t_j)-bad pseudo-trees in $\mathcal{T}_\varphi$ conditional on π. Notice that X_φ only depends on the random permutations $\Pi_{\psi,j}$ with $\psi \in \Psi$.

For each pseudo-tree $H \in \mathcal{T}'_\varphi$ at a $t_{j'}$-fan $\psi \in \Psi$, we already know what the image of t_i in H is. There are $c-1$ or c different images of T'' that could get matched to H at ψ, each having a distinct vertex as image of t_j. Thus, there are at least $c-1$ different vertices that could be the image of t_j in H. Since the permutation $\Pi_{\psi,j}$ is chosen uniformly at random, the probability that H will be part of a (t_i, t_j)-bad pseudo-tree is at most $\frac{1}{c-1}$. Now, by linearity of expectation,

$$\mathbb{E}[X_\varphi] \leqslant |\mathcal{T}_\varphi| \cdot \frac{1}{c-1} \leqslant \frac{c}{c-1}.$$

We will apply McDiarmid's inequality to the random variable Y_φ defined by $Y_\varphi := X_\varphi + c^{2/3}$. Clearly $\mathbb{E}[Y_\varphi] = \mathbb{E}[X_\varphi] + c^{2/3}$. Only the permutations $\Pi_{\psi,j}$ with $\psi \in \Psi$ affect X_φ and thus Y_φ. If two elements in one of these permutations are interchanged, then the structure of two pseudo-trees in $\mathcal{T}_\varphi$ changes. In particular, the number of (t_i, t_j)-bad trees in $\mathcal{T}_\varphi$ changes by at most 2. Thus, we can choose $d = 2$ in McDiarmid's inequality.

If $Y_\varphi \geqslant s$, then $X_\varphi \geqslant s - c^{2/3}$, and thus at least $s - c^{2/3}$ of the pseudo-trees in $\mathcal{T}_\varphi$ are (t_i, t_j)-bad. Let $H' \in \mathcal{T}'_\varphi$ be a part of a pseudo-tree H that is counted by X_φ. Let v_i and v_j denote the images of t_i and t_j in H. To verify that H is (t_i, t_j)-bad, we only need to know which edge in the j-blade of ψ gets mapped to the same position as the edges in H' in other blades of ψ. In other words, the vertex v_j is determined by the position of one element in the permutation $\Pi_{\psi,j}$, and thus $v_i = v_j$ can be certified by a single outcome. Thus, $X_\varphi \geqslant s - c^{2/3}$ can be certified by the outcomes of $s - c^{2/3} < s$ choices and we can choose $r = 1$ in McDiarmid's inequality.

By applying McDiarmid's inequality to Y_φ with $\lambda = \mathbb{E}[Y_\varphi]$, $d = 2$, $r = 1$, we get

$$\mathbb{P}\left[|Y_\varphi - \mathbb{E}[Y_\varphi]| > \mathbb{E}[Y_\varphi] + 120\sqrt{\mathbb{E}[Y_\varphi]}\right] \leqslant 4e^{-\frac{\mathbb{E}[Y_\varphi]}{32}} \leqslant 4e^{-\frac{c^{2/3}}{32}}.$$

Since $c \geqslant 10^9$ and $\mathbb{E}[Y_\varphi] \leqslant c^{2/3} + \frac{c}{c-1}$, we have $120\sqrt{\mathbb{E}[Y_\varphi]} < \frac{1}{2}\mathbb{E}[Y_\varphi]$ which implies

$$\mathbb{P}\left[X_\varphi > 2c^{2/3}\right] = \mathbb{P}\left[Y_\varphi > 3c^{2/3}\right] < \mathbb{P}\left[|Y_\varphi - \mathbb{E}[Y_\varphi]| > \frac{3}{2}\mathbb{E}[Y_\varphi]\right] \leqslant 4e^{-\frac{c^{2/3}}{32}}.$$

Now $\mathbb{P}[A_\varphi(t_i, t_j)|\pi] < 4e^{-c^{2/3}/32}$ and Claim 2A follows. □

Proof of Claim 2B:. Let $t_i \in V(T)$ be a fixed vertex different from t. Let P denote the path from t to t_i in T. Let t_j denote the second last vertex of P and let i denote the edge joining t_j and t_i. Now $T - i$ consists of two components, one of which contains t and t_j while the other one contains t_i. Let π be any fixed outcome of all permutations apart from those at the i-blades of t_j-fans. We show that the conditional probability $\mathbb{P}[B_\varphi(u|t_i)|\pi]$ is at most $4e^{-c^{2/3}/8}$. As π is arbitrary, this implies the general bound $\mathbb{P}[B_\varphi(u|t_i)] < 4e^{-c^{2/3}/8}$.

Let X_φ denote the random variable conditional on π which counts the number of pseudo-trees in $\mathcal{T}_\varphi$ where u is the image of t_i. The vertex u appears at most once in each t_j-fan, so by linearity of expectation we have

$$\mathbb{E}[X_\varphi] \leqslant |\mathcal{T}_\varphi| \cdot \frac{1}{c-1} \leqslant \frac{c}{c-1}\,.$$

We apply McDiarmid's inequality to the random variable $X_\varphi + c^{2/3}$. Swapping two positions in a permutation $\Pi_{\psi,i}$ can affect X_φ by at most 1 since u is incident to at most one edge of the i-blade of ψ. If $X_\varphi + c^{2/3} \geqslant s$, then this can be certified by revealing at most s positions in the random permutations. Thus, applying McDiarmid's inequality to the random variable $X_\varphi + c^{2/3}$ with $\lambda = \mathbb{E}[X_\varphi] + c^{2/3}$, $r = 1$, $d = 1$ yields

$$\mathbb{P}\left[X_\varphi > 2c^{2/3}\right] \leqslant 4e^{-c^{2/3}/8}\,.$$

Now $\mathbb{P}[B_\varphi(u|t_i)|\pi] < 4e^{-c^{2/3}/8}$ and Claim 2B follows. □

Now the proof of Claim 2 follows easily from Claims 2A and 2B.

Proof of Claim 2: By Claim 2A, we have

$$\mathbb{P}[A_\varphi] \leqslant \mathbb{P}\left[\bigcup_{\forall i<j} A_\varphi(t_i,t_j)\right] \leqslant \sum_{\forall i<j} \mathbb{P}\left[A_\varphi(t_i,t_j)\right] < 4m^2 e^{-c^{2/3}/32}\,.$$

Let $B_\varphi(u)$ be the event that the number of pseudo-trees in $\mathcal{T}_\varphi$ containing u is greater than $2mc^{2/3}$. Since u cannot be the image of t, we have, by Claim 2B,

$$\mathbb{P}[B_\varphi(u)] \leqslant \mathbb{P}\left[\bigcup_{\forall i} B_\varphi(u|t_i)\right] \leqslant \sum_{\forall i} \mathbb{P}[B_\varphi(u|t_i)] < 4me^{-c^{2/3}/8}\,.$$

Since each fan consists of at most cm edges, there are at most $cm + (cm)^2 + \ldots + (cm)^m$ vertices we can reach from φ via a path of length at most m. Thus, there are less than $2(cm)^m$ vertices u for which $\mathbb{P}[B_\varphi(u)]$ could be positive. In particular, we have

$$\mathbb{P}[B_\varphi] = \mathbb{P}\left[\bigcup_{\forall u,u\neq v} B_\varphi(u)\right] \leqslant \sum_{\forall u,u\neq v} \mathbb{P}[B_\varphi(u)] < 8(cm)^m me^{-c^{2/3}/8}$$

and Claim 2 follows from $\mathbb{P}[C_\varphi] \leqslant \mathbb{P}[A_\varphi] + \mathbb{P}[B_\varphi]$. □

This concludes the proof of Lemma 2.15. □

2.5 Repairing non-isomorphic copies

Let $\mathcal{H} \cup \mathcal{I}$ be the T-pseudo-decomposition given by Lemma 2.15. As described in Step 2 in Section 2, we use copies in $\mathcal{I}$ to repair the pseudo-trees in $\mathcal{H}$ that are not isomorphic to T. We apply Lemma 2.18 to show the existence of a suitable subset of $\mathcal{I}$ to perform the switches. The proof of Lemma 2.18 relies on the following probabilistic tools, see also [MR13]. The first one is a generalization of the Symmetric Local Lemma (Proposition 2.19).

PROPOSITION 2.21 Lovász Local Lemma

Let $A_1, ..., A_n$ be a finite set of events in some probability space Ω, and suppose that for some $J_i \subset [n]$, A_i is mutually independent of $\{A_j : j \notin J_i \cup \{i\}\}$. If there exist real numbers $x_1, ..., x_n$ in $(0,1)$ such that $\mathbb{P}[A_i] \leqslant x_i \prod_{j \in J_i}(1 - x_j)$ for every $i \in \{1, ..., n\}$, then $\mathbb{P}[\cap_{i=1}^n \overline{A_i}] > 0$.

PROPOSITION 2.22 Simple Concentration Bound

Let X be a random variable determined by n independent trials $T_1, ..., T_n$ such that changing the outcome of any one trial T_i can affect X by at most c. Then

$$\mathbb{P}[|X - \mathbb{E}[X]| > \lambda] \leqslant 2e^{-\lambda^2/(2c^2 n)}.$$

Proof of Lemma 2.18: Consider a pseudo-tree $H \in N_{\mathcal{H}}(v|t)$, and let $u \in V(H) \setminus \{v\}$. Since $\mathrm{conf}(\mathcal{H}'|t) \leqslant \delta$, there are no more than $\delta d_{\mathcal{H}'}(v|t)$ trees in $N_{\mathcal{H}'}(v|t)$ containing u. Thus, there are at least $(1 - \delta m) d_{\mathcal{H}'}(v|t)$ pseudo-copies of T in $N_{\mathcal{H}'}(v|t)$ that intersect H only in v. Since $d_{\mathcal{H}}(v|t) \leqslant \varepsilon d_{\mathcal{H}'}(v|t)$, we can associate a set $S(H)$ of $\lfloor \frac{1-\delta m}{\varepsilon} \rfloor$ pseudo-copies in $N_{\mathcal{H}'}(v|t)$ with each $H \in N_{\mathcal{H}}(v|t)$ such that each element of $N_{\mathcal{H}'}(v|t)$ is contained in at most one of these sets. We define the function f_t by choosing $f_t(H)$ uniformly at random from one of the pseudo-trees in $S(H)$. Clearly, any such function will satisfy the first two conditions of Lemma 2.18. All that remains to show is that with positive probability $d_{f_t(\mathcal{H})}(v|t') \leqslant 3\varepsilon d_{\mathcal{H}'}(v|t')$ holds for every compatible $v \in V(G), t' \in V(T)$.

The value of $d_{f_t(\mathcal{H})}(v|t')$ only depends on the set of pseudo-trees in $N_{\mathcal{H}'}(v|t')$ that are contained in some $S(H)$. Let $\mathcal{H}''$ be the collection of pseudo-copies of $\mathcal{H}'$ that are contained in some $S(H)$. Clearly, each tree in $N_{\mathcal{H}''}(v|t')$ can be matched with exactly one tree in $\mathcal{H}$ and this occurs with probability $\lfloor \frac{1-\delta m}{\varepsilon} \rfloor^{-1}$. By linearity of expectation,

$$\mathbb{E}[d_{f_t(\mathcal{H})}(v|t')] = \left\lfloor \frac{1 - \delta m}{\varepsilon} \right\rfloor^{-1} d_{\mathcal{H}''}(v|t') < 2\varepsilon d_{\mathcal{H}'}(v|t').$$

Let $A_{v,t'}$ be the event that $d_{f_t(\mathcal{H})}(v|t') > 3\varepsilon d_{\mathcal{H}'}(v|t')$. Note that $d_{f_t(\mathcal{H})}(v|t')$ is completely determined by at most $d_{\mathcal{H}''}(v|t')$ independent trials. Since the outcome of each trial can affect $d_{f_t(\mathcal{H})}(v|t')$ by at most 1, the Simple Concentration Bound gives

$$\mathbb{P}[A_{v,t'}] < 2e^{-\varepsilon^2 d_{\mathcal{H}''}(v|t')/2}.$$

We claim that $A_{v,t'}$ is mutually independent of all but at most $m\lfloor\frac{1-\delta m}{\varepsilon}\rfloor d_{\mathcal{H}''}(v|t')$ other events $A_{v',t''}$. Indeed, $A_{v,t'}$ depends on at most $d_{\mathcal{H}''}(v|t')$ random trials, and in each trial we have a choice of $\lfloor\frac{1-\delta m}{\varepsilon}\rfloor$ trees to match. Each tree affects precisely m events other than $A_{v,t'}$.

Now we apply Proposition 2.21 to show that with positive probability none of the events $A_{v,t'}$ occur. Set $x = \frac{\varepsilon^4}{8}$. It is sufficient to show that

$$x\,(1-x)^{m\lfloor\frac{1-\delta m}{\varepsilon}\rfloor d_{\mathcal{H}''}(v|t')} \geqslant \mathbb{P}[A_{v,t'}]$$

holds for all compatible $v \in V(G)$, $t' \in V(T)$. If $d_{\mathcal{H}''}(v|t') < \left(\frac{2}{\varepsilon}\right)^6$, then $d_{f_t(\mathcal{H})}(v|t') < \left(\frac{2}{\varepsilon}\right)^6 < 3\varepsilon d_{\mathcal{H}'}(v|t')$, so $\mathbb{P}[A_{v,t'}] = 0$. If $d_{\mathcal{H}''}(v|t') \geqslant \left(\frac{2}{\varepsilon}\right)^6$, then we have

$$\begin{aligned}
x\,(1-x)^{m\lfloor\frac{1-\delta m}{\varepsilon}\rfloor d_{\mathcal{H}''}(v|t')} &\geqslant x\,(1-x)^{d_{\mathcal{H}''}(v|t')/\varepsilon^2} \\
&\geqslant xe^{-2xd_{\mathcal{H}''}(v|t')/\varepsilon^2} \\
&\geqslant \mathbb{P}[A_{v,t'}] \cdot \frac{x}{2} \cdot e^{\frac{\varepsilon^2}{4} d_{\mathcal{H}''}(v|t')} \\
&\geqslant \mathbb{P}[A_{v,t'}] \cdot \left(\frac{\varepsilon}{2}\right)^6 d_{\mathcal{H}''}(v|t') \\
&\geqslant \mathbb{P}[A_{v,t'}]\,.
\end{aligned}$$

By the Local Lemma, there is a positive probability that none of the bad events occur. Thus, there exists a function f_t with the desired properties. □

We now have all ingredients for the proof of Theorem 2.12.

Proof of Theorem 2.12: As described in Step 2 of Section 2, let $t_0, \dots, t_m$ be a labelling of the vertices of T such that $T[t_0, \dots, t_i]$ is connected for every $i \in \{1, \dots m\}$. We also label the edges of T so that e_i denotes the edge joining t_i with $T[t_0, \dots, t_{i-1}]$ for every $i \in \{1, \dots m\}$. Set $\varepsilon_i = 5^{i-m}/15m$ for $i \in \{1, \dots, m\}$. We are going to construct a sequence $(\mathcal{H}_i \cup \mathcal{I}_i)_{i=1}^m$ of T-pseudo-decompositions of G such that the following holds:

- $\mathcal{I}_i$ is a collection of isomorphic copies of T for every $i \in \{1, \dots m\}$;
- $\mathcal{H}_i$ is i-good for every $i \in \{1, \dots m\}$;
- $d_{\mathcal{I}_i}(v|t) > \max\{22/\varepsilon_i^7, d_{\mathcal{H}_i}(v|t)/\varepsilon_i\}$ for every compatible $v \in V(G)$, $t \in V(T)$;

- $\mathrm{conf}(\mathcal{I}_i) \leqslant \varepsilon_i$ for every $i \in \{1, \dots, m\}$.

Since the minimum degree in each color in G is at least 10^{50m}, we can apply Lemma 2.15 with parameters $\varepsilon = \delta = 10^{-2m}$. Let $\mathcal{H} \cup \mathcal{I}$ denote the resulting T-pseudo-decomposition. Clearly $\mathcal{H} \cup \mathcal{I}$ satisfies the conditions for $\mathcal{H}_1 \cup \mathcal{I}_1$. Let $i \in \{2, \dots, m\}$ and suppose we have constructed $\mathcal{H}_{i-1} \cup \mathcal{I}_{i-1}$ such that the conditions above are satisfied. We need to repair the pseudo-trees in $\mathcal{H}_{i-1}$ that are not i-good. Since the pseudo-trees in $\mathcal{H}_{i-1}$ are all $(i-1)$-good, we can achieve this by making i-switches. Let t_j be the endpoint of e_i that is different from t_i. Let $f_j : \mathcal{H}_{i-1} \to \mathcal{I}_{i-1}$ be the function we get by applying Lemma 2.18 with $\mathcal{H} = \mathcal{H}_{i-1}$, $\mathcal{H}' = \mathcal{I}_{i-1}$, $\varepsilon = \delta = \varepsilon_{i-1}$, and $t = t_j$. Now $f_j(\mathcal{H}_{i-1})$ is the set of trees we use to repair the pseudo-trees in $\mathcal{H}_{i-1}$ that are not i-good. Set

$$\begin{aligned} \mathcal{H}_i &= \bigcup_{H \in \mathcal{H}_{i-1}} \mathrm{sw}_i(H, f_j(H)) \text{ and} \\ \mathcal{I}_i &= \mathcal{I}_{i-1} \setminus f_j(\mathcal{H}_{i-1}) \,, \end{aligned}$$

where $\mathrm{sw}_i(H, f_j(H))$ denotes the i-switch of H and $f_j(H)$ as defined in Section 2. Since $H \cap f_j(H) = \{v\}$ for every $H \in N_{\mathcal{H}_{i-1}}(v|t_j)$, the two pseudo-copies in $\mathrm{sw}_i(H, f_j(H))$ are both i-good.

Notice that the degree $d_{\mathcal{H}_i}(v|t)$ of a vertex is invariant under i-switches between pseudo-trees in $\mathcal{H}_i$. Since $d_{f_j(\mathcal{H}_{i-1})}(v|t) \leqslant 3\varepsilon_{i-1} d_{\mathcal{I}_{i-1}}(v|t)$ holds for compatible $v \in V(G)$ and $t \in V(T)$, we have $d_{\mathcal{I}_i}(v|t) \geqslant (1 - 3\varepsilon_{i-1}) d_{\mathcal{I}_{i-1}}(v|t)$ and $d_{\mathcal{H}_i}(v|t) \leqslant 4\varepsilon_{i-1} d_{\mathcal{I}_{i-1}}(v|t)$. Thus,

$$d_{\mathcal{H}_i}(v|t) \leqslant \frac{4\varepsilon_{i-1}}{(1 - 3\varepsilon_{i-1})} d_{\mathcal{I}_i}(v|t) \leqslant 5\varepsilon_{i-1} d_{\mathcal{I}_i}(v|t) = \varepsilon_i d_{\mathcal{I}_i}(v|t) \,,$$

$$d_{\mathcal{I}_i}(v|t) \geqslant (1 - 3\varepsilon_{i-1}) d_{\mathcal{I}_{i-1}}(v|t) \geqslant 22 \frac{1 - 3\varepsilon_{i-1}}{\varepsilon_{i-1}^7} \geqslant \frac{22}{\varepsilon_i^7}$$

and

$$\mathrm{conf}(\mathcal{I}_i) \leqslant \frac{\mathrm{conf}(\mathcal{I}_{i-1})}{1 - 3\varepsilon_{i-1}} \leqslant \frac{5}{4} \varepsilon_{i-1} < \varepsilon_i \,.$$

Hence, the T-pseudo-decomposition $\mathcal{H}_i \cup \mathcal{I}_i$ has the desired properties. In particular, $\mathcal{H}_m$ is m-good and $\mathcal{H}_m \cup \mathcal{I}_m$ is a T-decomposition of G. □

Chapter 3

Beyond the Barát-Thomassen Conjecture

3.1 Finding a weaker hypothesis

While the Tutte's 3-flow conjecture states that the required amount of edge-connectivity is 4, Thomassen [Tho12] could only achieve 8. One year later, Lovász et al [LTWZ13] showed that 6-edge-connectivity is sufficient, and so cut half the distance between the state-of-the-art and the 3-Flow Conjecture. The difference between 4-edge-connectivity and 6-edge-connectivity is indeed not as narrow as it may sound. The edge-connectivity is an expensive condition since increasing the amount of edge-connectivity on a graph by 1 means another layer of global connection needs to be added, while global connections are hard to achieve in large graphs.

Unfortunately, our proof of the Barát-Thomassen conjecture in Chapter 2 provides a terrible exponential bound, $10^{O(|E(T)|)}$, for the amount of edge-connectivity of G due to the probabilistic approach. To reduce this amount to a reasonable bound, one may need a completely new approach, which is left for future work. On the other hand, given the high cost of edge-connectivity, it would be reasonable to some cheaper conditions for the Barát-Thomassen conjecture rather than the edge-connectivity.

The content of this chapter is covered in paper *Edge-partitioning a graph into paths: beyond the barát-thomassen conjecture*, Julien Bensmail, Ararat Harutyunyan, Tien-Nam Le, and Stéphan Thomassé, *Combinatorica*, pages 1–25, 2018.

QUESTION 3.1

> *Is it possible to replace the high edge-connectivity hypothesis in the Barát-Thomassen conjecture by some weaker ones which do not require such strong global connections?*

One prominent candidate is high minimum degree, a much less costly requirement, since roughly speaking it only requires that every vertex has enough arbitrary connections to the rest of the large graph rather than every vertex must have high global connections. However, similar to the Tutte's flow conjectures, it is easy to see that in order to the Barát-Thomassen conjecture holds, the graph G must have some minimum amount of edge-connectivity. One simple example is where G is the disjoint union of two cliques of size $4n + 2$ for an arbitrarily large n. Then G has very high minimum degree but cannot be decomposed into copies of P_2, the path of length 2. The reason is that each clique of G has odd number of edges, so there is always one edge left out in each clique which could not be joint to become a copy of P_2. A more complex example in Proposition 3.8 and Figure 3.1 gives a counterexample of a 2-edge-connected graph G with very large minimum degree, which could not have a P_9-decomposition.

The main result of this chapter is the following theorem, which states that if G is 24-edge-connected, then high minimum degree is a sufficient condition so that the Barát-Thomassen conjecture holds when the pattern is a path.

THEOREM 3.2

> *For every integer $\ell \geqslant 2$, there exists d_ℓ such that every 24-edge-connected graph G with minimum degree at least d_ℓ has a decomposition into paths of length ℓ and a path of length at most ℓ.*

Let us again emphasize that the main point in the statement of Theorem 3.2 is that the required edge-connectivity, namely 24, is constant and not dependent on the path length ℓ as in the statement of the Barát-Thomassen conjecture. Nevertheless, this theorem still leaves a gap to the lower bound the amount of edge-connectivity. Very recently, Klimošová and Thomassé drastically improved the bound from 24 to 3 [KT17], and so by combining with Proposition 3.8 to provide a tight bound for path patterns in general. Nonetheless, the true bound for specific patterns has not yet been reached. For example, one may ask what is the minimum amount of edge-connectivity required when the pattern is P_3 or P_5, which is still an open question.

It is natural to expect that we can similarly replace high edge-connectivity by high minimum degree when the pattern is a tree. However, it was proved, also in [KT17], that this does not hold, and so Question 3.1 is still open when the pattern is a tree.

Some notations and preliminary results

In Chapter 2, we used "collection" as a set of copies of the pattern. In this chapter we will develop a system of notions around the term *collection*. The proof contains two main parts. We will show in Section 3.2 that the graph G contains a large collection with low conflict and covering almost entire G, and then then in Section 3.3 that G contains another sparse collection, which has a tree-shape and spanning all over G. These two collections will complement each other to return a path-decomposition of G without any conflict, and so proves Theorems 3.2.

Given a graph G and a subset X of $V(G)$, we denote by $d_X(v)$ the degree of v in the subgraph of G induced by $X \cup \{v\}$. Given two graphs $G = (V, E)$ and $H = (V, F)$ with $F \subseteq E$, we denote by $G \backslash H$ the graph $(V, E \backslash F)$.

We now introduce some notions about the density of a subgraph with respect to the original graph. Let $H = (V, F)$ be a spanning subgraph of a graph $G = (V, E)$. Let α be some real number in $[0, 1]$. We say that H is *α-sparse* in G if $d_H(v) \leqslant \alpha d_G(v)$ for all vertices v of G. Conversely, we say that H is *α-dense* in G if $d_H(v) \geqslant \alpha d_G(v)$ for all vertices v of G. We will also heavily depend on subgraphs of G which are both (roughly) α-sparse and α-dense. This definition depends on the length of the path ℓ. We say that H is an *α-fraction* of G if $\alpha d_G(v) - 10\ell^\ell \leqslant d_H(v) \leqslant \alpha d_G(v) + 10\ell^\ell$.

Given an (improper) edge-coloring ϕ of some graph G and a color i, for every vertex v of G we denote by $d_i(v)$ the number of i-colored edges incident to v. We call ϕ *nearly equitable* if, for every vertex v and every pair of colors $i \neq j$, we have $|d_i(v) - d_j(v)| \leqslant 2$. We can now recall a result of de Werra (cf. [SSTF12], Theorem 8.7), and its corollary concerning $1/k$-fractions.

PROPOSITION 3.3

Let $k \geqslant 1$. Every graph has a nearly equitable improper k-edge-coloring.

PROPOSITION 3.4

Let $k \geqslant 1$. Every graph $G = (V, E)$ has a spanning subgraph $H = (V, F)$ such that $|d_H(v) - d_G(v)/k| \leqslant 2$ for every vertex v.

We now recall three results on oriented graphs. The first of these is a folklore result on balanced orientations of graphs.

PROPOSITION 3.5

Every multigraph G has an orientation D such that $|d_D^-(v) - d_D^+(v)| \leqslant 1$ for every vertex v.

The proof is straightforward. We first arbitrarily pair vertices of odd degree of G,

then add a dummy edge between every pair to obtain a multigraph G' in which every vertex has even degree. Orienting the edges of G' as they are encountered when going along an Eulerian tour, we then deduce an orientation D' of G' such that $d^-_{D'}(v) = d^+_{D'}(v)$ for every vertex v. Removing the dummy edges results in a desired orientation D of G.

The second result is a result of Nash-Williams (see [NW60]) implying that any graph with large edge-connectivity admits a balanced orientation with large arc-connectivity. In the following, a digraph D is *k-arc-strong* if the removal of any set of at most $k-1$ arcs leaves D strongly-connected.

PROPOSITION 3.6

Every $2k$-edge-connected multigraph has an orientation D such that D is k-arc-strong and such that $|d^-(v) - d^+(v)| \leqslant 1$ for every vertex v.

The third result we recall is due to Edmonds (see [Edm73]) and expresses a condition for a digraph to admit many arc-disjoint rooted arborescences. In the statement, an *out-arborescence* of a digraph D refers to a rooted spanning tree T of D whose arcs are oriented in such a way that the root has in-degree 0, and every other vertex has in-degree 1.

PROPOSITION 3.7

A directed multigraph with a special vertex z has k arc-disjoint out-arborescences rooted at z if and only if the number of arc-disjoint paths between z and any vertex is at least k.

For the sake of completeness, we end this section by proving the claim that 3-edge-connectivity is necessary when the pattern is P_9.

PROPOSITION 3.8

There exist 2-edge-connected graphs with arbitrarily large minimum degree admitting no P_9-decomposition.

Proof. We start from the 2-edge-connected graph G depicted in Figure 3.1, which admits no P_9-decomposition. To obtain a 2-edge-connected graph with arbitrarily large minimum degree d from it, just consider any 2-edge-connected graph G' with sufficiently large minimum degree (i.e., at least d) and verifying $|E(G')| \equiv 7 \pmod 9$. Then consider any vertex v of G with small degree, and add two edges from v to a new copy of G'. Repeating this transformation as long as necessary, we get a new graph which is still 2-edge-connected, with minimum degree at least d and whose size is a multiple of 9 (due to the size of G and G'), but with no P_9-decomposition – otherwise, it can be easily checked that G would admit a P_9-decomposition, a contradiction. □

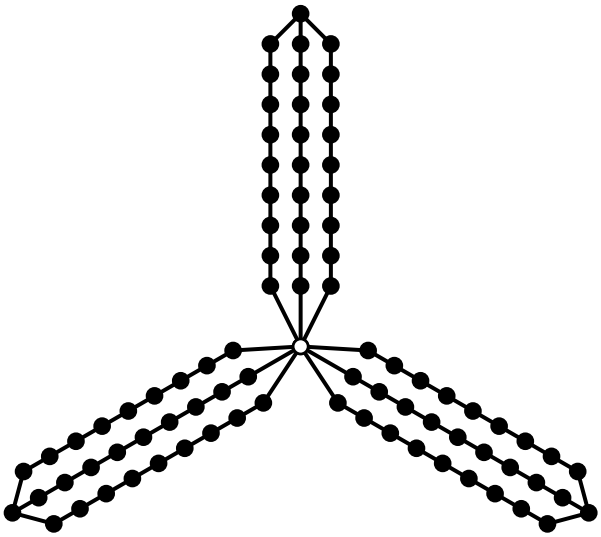

FIGURE 3.1 – Part of the construction for obtaining 2-edge-connected graphs with arbitrarily large minimum degree but no P_9-decomposition.

3.2 Path-collections

Fix P_ℓ as our pattern and let $G = (V, E)$ be a graph. A *path-collection* $\mathcal{P}$ *on* G is a set of edge-disjoint paths of G. We denote by $U_\mathcal{P} = (V, E')$ the graph where E' is the set of edges of paths in $\mathcal{P}$. If $U_\mathcal{P} = G$ then $\mathcal{P}$ is said to be a *path-decomposition* of G. For convenience, from now on, we say *collection* for path-collection and *decomposition* for path-decomposition.

Let us denote by $H_\mathcal{P} = (V, E'')$ the multigraph where each edge $uv \in E''$ corresponds to a path between u and v in $\mathcal{P}$ (if $\mathcal{P}$ contains several paths from u to v, we have as many edges $uv \in E''$). The *degree* of a vertex v in $\mathcal{P}$, denoted $d_\mathcal{P}(v)$, is the degree (with multiplicity) of v in $H_\mathcal{P}$, which is also the number of paths in $\mathcal{P}$ with endpoint v.

We need also to speak of the lengths of the paths in $\mathcal{P}$. Let us say that $\mathcal{P}$ is an *ℓ-collection* if all paths in $\mathcal{P}$ have length ℓ, a *$(\geqslant \ell)$-collection* if all paths in $\mathcal{P}$ have length at least ℓ, an *$(\ell_1, \ell_2, ...)$-collection* if all paths in $\mathcal{P}$ have length among $\{\ell_1, \ell_2, ...\}$, and an *$[\ell, \ell + i]$-collection* if all paths in $\mathcal{P}$ have length in the interval $[\ell, \ell + i]$.

Similar to Chapter 2, we also use notions of conflict and conflict ratio in this proof. Two edge-disjoint paths of G sharing an endpoint v are *conflicting* if they also intersect at some vertex different from v. Equivalently, we say that two paths of $\mathcal{P}$ issued from the same vertex are *conflicting* if the corresponding paths in $U_\mathcal{P}$ are conflicting. In general, the paths of a collection can pairwise intersect, and hence we would like to measure how much. For every vertex v, let $\mathcal{P}(v)$ be the set of paths in $\mathcal{P}$ *incident* with

v, i.e., starting or ending at v. The *conflict ratio* of v is

$$\mathrm{conf}_{\mathcal{P}}(v) := \frac{\max_{w \neq v} \left|\{P \in \mathcal{P}(v) : w \in P\}\right|}{d_{\mathcal{P}}(v)}.$$

We denote the *conflict ratio* of $\mathcal{P}$ by $\mathrm{conf}(\mathcal{P}) := \max_v \mathrm{conf}_{\mathcal{P}}(v)$. We always have $\mathrm{conf}(\mathcal{P}) \leqslant 1$ since $|\mathcal{P}(v)| = d_{\mathcal{P}}(v)$.

We now prove that every graph with large enough minimum degree contains a $(\geqslant \ell)$-collection meeting particular properties: low conflict, well-balanced, and covers almost entire G. Note that in this section, the edge-connectivity of G will not be utilized.

THEOREM 3.9

Let ℓ be a positive integer, and ε be an arbitrarily small positive real number. There exists L such that if $G = (V, E)$ is a graph with minimum degree at least L, then there is an ℓ-collection $\mathcal{P}$ on G with

- $\mathrm{conf}(\mathcal{P}) \leqslant \varepsilon$,
- $d_{\mathcal{P}}(v)/d_G(v) \in \left[\frac{1-\varepsilon}{\ell}, \frac{1+\varepsilon}{\ell}\right]$ *for every vertex v, and*
- $d_{G \setminus U_{\mathcal{P}}}(v) \leqslant \varepsilon d_{\mathcal{P}}(v)$ *for every vertex v.*

Proof. Let $c := [\sqrt{L}]$ and $b := [c^{2/3}]$, and pick L so that $b \gg \ell$. According to Proposition 3.3, we can nearly equitably color the edges of G with ℓ colors. For every color i, applying Proposition 3.5 we can orient the i-colored edges so that the numbers of in-edges and out-edges of color i incident to every vertex v differ by at most 1. Let $E_i^-(v)$ and $E_i^+(v)$ be the sets of i-colored in-edges and out-edges, respectively, incident to v. Then, for every color $i \in \{1, ..., \ell - 1\}$, we have

$$\left||E_i^-(v)| - |E_{i+1}^+(v)|\right| \leqslant 2.$$

For the sake of convenience, we would like to have that $|E_i^-(v)| = |E_{i+1}^+(v)|$ for all i and v. To this end, we add a dummy vertex v_0 to G. Now, if $|E_i^-(v)| - |E_{i+1}^+(v)| = k > 0$, then we add k dummy edges of color $i+1$ from v to v_0 to equalize $|E_i^-(v)|$ and $|E_{i+1}^+(v)|$. Similarly, if $|E_{i+1}^+(v)| - |E_i^-(v)| = k > 0$, then we add k dummy edges of color i from v_0 to v.

Now, for every $v \in V(G)$ and color $i \in \{1, ..., \ell\}$, we choose $r_{v,i} \in \{0, \ldots, c-2\}$ such that $E_i^-(v) \equiv r_{v,i} \pmod{c-1}$. Since the minimum degree in each color in G is greater than $c(c-2)$, we can partition every set $E_i^-(v)$ into subsets of size c and $c-1$ so that precisely $r_{v,i}$ of them have size c. As $E_{i+1}^+(v) = E_i^-(v)$, we can similarly partition every set $E_{i+1}^+(v)$ into subsets of size c and $c-1$ so that precisely $r_{v,i}$ of them have size c.

We call these subsets of edges *i-half-cones* and *$(i+1)$-half-cones*, respectively. Now, for each vertex v and color i, $1 \leqslant i \leqslant \ell - 1$, we arbitrarily pair i-half-cones of $E_i^-(v)$ with $(i+1)$-half-cones $E_{i+1}^+(v)$ in a way such that in each pair the size of the two half-cones are equal. We call such a pair an *i-cone* at vertex v. Thus, an i-cone φ at some vertex v consists of an i-half-cone φ^- and an $(i+1)$-half-cone φ^+ with $|\varphi^-| = |\varphi^+|$. Note that an edge e of color i directed from a vertex u to a vertex v in G appears both in an i-half-cone of $E_i^+(u)$ as well as in an i-half-cone of $E_i^-(v)$, but we do not require these two i-half-cones to have the same size. By convention, we do not create a cone at the dummy vertex v_0. However, each edge uv_0 will still be inside a cone at vertex u. We also remark that the 1-half-cones of $E_1^+(v)$ and the ℓ-half-cones of $E_\ell^-(v)$ do not get paired with other half-cones. Nevertheless, we will adopt the convention that whenever we talk of a general cone φ, we will assume that φ might also consist of a single 1-half-cone or ℓ-half-cone of the aforementioned type.

We now have a fixed set of cones on G. To obtain our desired collection, we will use the cone structure to construct rainbow paths of length ℓ, i.e., paths where for all i the i^{th} edge of every path is of color i. One way to obtain this is to randomly match edges of the two half-cones of every cone. Indeed, this is what we do. For each cone φ we carry out random permutations π_φ^- of the edges of φ^- and π_φ^+ of the edges of φ^+. We then pair the edges $\pi_\varphi^-(k)$ and $\pi_\varphi^+(k)$ for each $1 \leqslant k \leqslant c$. If φ is actually a special 1-half-cone or ℓ-half-cone, then there is only one random permutation performed at φ, which will have no effect on the decomposition as will be apparent shortly. Note that each edge $e = uv$ of G, with the exception of some edges of 1-cones, some edges of ℓ-cones and the dummy edges, is in exactly two cones - one centered at u and the second centered at v. Thus, e is involved in two random permutations corresponding to the two permutations of the two half-cones containing it. Therefore, given the random matchings, each non-dummy edge $e = uv$ of color i, $1 < i < \ell$, is paired exactly with one edge of color $i-1$ (which enters u) and one edge of color $i+1$ (which exits v). From an arbitrary edge, we can thus go forward and backward by edges paired with it until we reach edges of color ℓ or 1 (unless we reach dummy edges). Thus, the random matchings yield a natural decomposition of all edges of G into edge-disjoint *walks*. Unfortunately, some of the walks will not be paths. We will divide the walks into three types. Of the first type are those walks which are paths, and thus by construction they are necessarily isomorphic to P_ℓ. A walk that is not a path and which does not use the dummy vertex v_0 is called a *bad* walk; note that every bad walk is of length ℓ. A walk that uses the dummy vertex v_0 is called a *short* walk. Note that a short walk is no longer extended from v_0 as there is no cone centered at v_0.

For each cone φ, there are $c-1$ or c walks via φ, depending on $|\varphi|$. We will show that, with high probability, the number of bad or short walks via φ is negligible compared to c. We then will argue that proving this statement for all the cones is sufficient for us to extract a dense collection from G.

Denote $P_\ell := x_0x_1...x_\ell$. We first focus on bad walks. Suppose that φ is a k-cone

at some vertex v, and i, j are two colors. We say that a bad walk $P = u_0u_1...u_\ell$ going through φ is *(i,j)-bad* if its i^{th} vertex and j^{th} vertex are the same, that is, $u_j = u_i$. Let $A_\varphi(i,j)$ be the event that the number of (i,j)-bad walks going through the cone φ is greater than b. We will show that $\mathbb{P}[A_\varphi(i,j)] < 4e^{-c^{2/3}/64}$.

Denote by $P_{i,k}$ and $P_{j,k}$ the subpaths from x_i to x_k, and x_j to x_k in P_ℓ, respectively. In case one of these paths is contained in another, we may assume that $P_{i,k}$ is contained in $P_{j,k}$. Let $x_{j'}$ be the neighbor of x_j in $P_{j,k}$. Note that $j' \in \{j-1, j+1\}$. Let $\mathcal{P}_\varphi$ be the set of walks that go through φ which are not short. Clearly, $|\mathcal{P}_\varphi| \leqslant c$.

We define $\Omega_{j'}$ to be the set of all j'-cones in G if $j' = j-1$, and the set of all j-cones if $j' = j+1$. Let Π be an arbitrary but fixed outcome of all permutations at all cones except the set of permutations on $\Omega_{j'}$. In other words, given Π, we only need to know the outcomes of the set of permutations $\{\pi^+_\varphi, \pi^-_\varphi \mid \varphi \in \Omega_{j'}\}$ to know the decomposition of the walks in G. We will condition on Π; that is, we will show that $\mathbb{P}[A_\varphi(i,j)] \mid \Pi] < 4e^{-c^{2/3}/64}$ for any Π. Clearly, since Π is arbitrary, this is sufficient to give us the uniform bound $\mathbb{P}[A_\varphi(i,j)] < 4e^{-c^{2/3}/64}$.

Let $\mathcal{P}'_\varphi$ denote the set of walks $\mathcal{P}_\varphi$ conditional on Π. Let X_φ be the number of (i,j)-bad walks going through the cone φ conditional on Π. By fixing Π, the set $\mathcal{P}'_\varphi$ is also fixed. Indeed, each $P' \in \mathcal{P}'_\varphi$ is a partial subwalk, where we know the vertex of P' that lies in some half-cone of a cone $\psi \in \Omega_{j'}$. Note that the vertex u_i of P' corresponding to x_i is already known. Moreover, the vertex $u_{j'}$ corresponding to the vertex $x_{j'}$ is known as well.

Note that whether P' is (i,j)-bad depends only on the permutations π^-_ψ and π^+_ψ. Note that there are $c-1$ or c different images possible to match $u_{j'}$ when the random permutations π^-_ψ and π^+_ψ are carried out, and only one of which could possibly be u_i. Thus, the probability that P' is (i,j)-bad is at most $\frac{1}{c-1}$.

Now, by linearity of expectation,

$$\mathbb{E}[X_\varphi] \leqslant |\mathcal{P}_\varphi| \cdot \frac{1}{c-1} \leqslant \frac{c}{c-1}.$$

We will apply McDiarmid's inequality to the random variable Y_φ defined by $Y_\varphi := X_\varphi + c^{2/3}$. Clearly $\mathbb{E}[Y_\varphi] = \mathbb{E}[X_\varphi] + c^{2/3} \in [c^{2/3}, c^{2/3}+2]$. Only the permutations π^-_ψ, π^+_ψ with $\psi \in \Omega_{j'}$ affect X_φ and thus Y_φ. If two elements in one of these permutations are interchanged, then the structure of two walks in $\mathcal{P}_\varphi$ changes. However, clearly the number of (i,j)-bad walks in $\mathcal{P}_\varphi$ cannot change by more than 2. Thus, we can choose $d = 2$ in McDiarmid's inequality.

If $Y_\varphi \geqslant s$, then $X_\varphi \geqslant s - c^{2/3}$, and thus at least $s - c^{2/3}$ of the walks in $\mathcal{P}_\varphi$ are (i,j)-bad. Let $P' \in \mathcal{P}'_\varphi$ be a subwalk of a walk P that is counted by X_φ. As before, let $u_i = u_j$ denote the images of x_i and x_j in P, and $\psi \in \Omega_{j'}$ the cone through which P' passes. To verify that P is (i,j)-bad, we only need to reveal the two elements $\pi^+_\psi(s), \pi^-_\psi(s)$, where $1 \leqslant s \leqslant c$ is the value such that the edge $u_{j'}u_j \in \{\pi^+_\psi(s), \pi^-_\psi(s)\}$.

Thus, $X_\varphi \geqslant s - c^{2/3}$ can be certified by the outcomes of $2(s - c^{2/3}) < 2s$ choices and we can choose $r = 2$ in McDiarmid's inequality. By applying McDiarmid's inequality to Y_φ with $\lambda = \mathbb{E}[Y_\varphi]$, $d = 2$, $r = 2$, we get

$$\mathbb{P}\left[|Y_\varphi - \mathbb{E}[Y_\varphi]| > \mathbb{E}[Y_\varphi] + 120\sqrt{2\mathbb{E}[Y_\varphi]}\right] \leqslant 4e^{-\frac{\mathbb{E}[Y_\varphi]}{64}} \leqslant 4e^{-\frac{c^{2/3}}{64}}$$

and thus $\mathbb{P}\left[X_\varphi > 2c^{2/3}\right] \leqslant 4e^{-c^{2/3}/64}$. So we have $\mathbb{P}[A_\varphi(i,j)|\Pi] < 4e^{-c^{2/3}/64}$. Since Π is arbitrary it follows that $\mathbb{P}[A_\varphi(i,j)] < 4e^{-c^{2/3}/64}$. Let A_φ be the event that there are more than $\ell^2 b$ bad walks via φ. Then

$$\mathbb{P}[A_\varphi] \leqslant \mathbb{P}\Big[\bigcup_{\forall i,j} A_\varphi(i,j)\Big] \leqslant \sum_{\forall i,j} \mathbb{P}[A_\varphi(i,j)] < 4\ell^2 e^{-c^{2/3}/64}.$$

We still consider the same cone φ. For an integer $j \neq k$ and vertex u, let $B_\varphi(j,u)$ be the event that the number of walks via φ, which maps x_j to u, is greater than b, and let $B_\varphi(u)$ be the event that the number of walks of φ containing u is greater than ℓb.

We show that $\mathbb{P}[B_\varphi(j,u)] < 4e^{-c^{2/3}/64}$. As the computation is virtually identical to the case of $\mathbb{P}[A_\varphi(i,j)]$, we only highlight the differences. As before, let $x_{j'}$ be the vertex adjacent to x_j on $P_{j,k}$, and let Π be an arbitrary but fixed outcome of all permutations at all cones except the set of permutations on $\Omega_{j'}$. It suffices to show that $\mathbb{P}[B_\varphi(j,u) \mid \Pi] < 4e^{-c^{2/3}/64}$.

Let X_φ denote the random variable conditional on Π which counts the number of walks in $\mathcal{P}_\varphi$ where u is the image of x_j. The vertex u appears at most once in each cone of $\Omega_{j'}$, so by linearity of expectation we have

$$\mathbb{E}[X_\varphi] \leqslant |\mathcal{P}_\varphi| \cdot \frac{1}{c-1} \leqslant \frac{c}{c-1}.$$

We again apply McDiarmid's inequality to the random variable Y_φ defined by $Y_\varphi := X_\varphi + c^{2/3}$. As before, $\mathbb{E}[Y_\varphi] = \mathbb{E}[X_\varphi] + c^{2/3}$.

Since the vertex u appears at most once in each cone of $\Omega_{j'}$, swapping two positions in any permutation of a half-cone in $\Omega_{j'}$ can affect X_φ by at most 1. Thus, we can choose $d = 1$ in McDiarmid's inequality.

If $Y_\varphi \geqslant s$, then $X_\varphi \geqslant s - c^{2/3}$. Let P' be a subwalk that is counted by X_φ. As before, we can certify that P' is counted by X_φ by considering only $\psi \in \Omega_{j'}$, the cone through which P' passes.

To certify that P' is counted by X_φ we only need to reveal the two elements $\pi^+_\psi(s), \pi^-_\psi(s)$, where s is the value such that one of the edges $\pi^+_\psi(s), \pi^-_\psi(s)$ contains the endpoint u. Thus, $X_\varphi \geqslant s - c^{2/3}$ can be certified by the outcomes of $2(s - c^{2/3}) < 2s$ choices and we can choose $r = 2$ in McDiarmid's inequality. Thus, by a similar argument as above we obtain that $\mathbb{P}[B_\varphi(j,u)] < 4e^{-c^{2/3}/64}$. Now,

$$\mathbb{P}\left[B_\varphi(u)\right] \leqslant \mathbb{P}\Big[\bigcup_{\forall i} B_\varphi(i,u)\Big] \leqslant \sum_{\forall i} \mathbb{P}\left[B_\varphi(i,u)\right] < 4\ell e^{-c^{2/3}/64}.$$

Let B_φ be the event that there exists a vertex u such that more than ℓb walks of φ contain u. The number of vertices u that could possibly appear in the walks $\mathcal{P}_\varphi$ is at most $c + c^2 + ... + c^\ell < c^{\ell+1}$. Hence,

$$\mathbb{P}[B_\varphi] = \mathbb{P}\Big[\bigcup_{\forall u} B_\varphi(u)\Big] \leqslant \sum_{\forall u} \mathbb{P}[B_\varphi(u)] < 4c^{\ell+1}\ell e^{-c^{2/3}/64}.$$

Let $B'_\varphi(j)$ be the event that the number of walks via φ such that they enter v_0 at exactly their j^{th}-vertex is greater than b, and let B'_φ be the event that the number of walks of φ containing v_0 is greater than ℓb. We upper bound $\mathbb{P}[B'_\varphi(j)]$.

The argument is virtually identical to that of the estimate above. We apply McDiarmid's inequality to the random variable $Y_\varphi := X_\varphi + c^{2/3}$, where X_φ is the number of walks via φ that enter v_0 at the j^{th} edge conditional on Π. As before, we obtain that $\mathbb{E}[X_\varphi] \leqslant c/(c-1)$, $d = 1, r = 2$, yielding $\mathbb{P}[B'_\varphi(j)] \leqslant 4e^{-c^{2/3}/64}$. Thus, $\mathbb{P}[B'_\varphi] < 4\ell e^{-c^{2/3}/64}$.

Let $J_\varphi = A_\varphi \cup B_\varphi \cup B'_\varphi$. Then

$$\mathbb{P}[J_\varphi] \leqslant \mathbb{P}[A_\varphi] + \mathbb{P}[B_\varphi] + \mathbb{P}[B'_\varphi] < (\ell^2 + c^{\ell+1}\ell + \ell)4e^{-c^{2/3}/64} < e^{-b/100}.$$

Let $\mathcal{J}_\varphi$ be the set of events J_ψ that are not mutually independent of J_φ. Note that the number of permutations determining J_φ is at most $(2c) + (2c)^2 + ... + (2c)^\ell < c^{\ell+1}$. Indeed, $c^{\ell+1}$ is an upper bound on the number of walks of length ℓ that could contain an edge of φ. Each such permutation itself could affect at most $c + ... + c^\ell < c^{\ell+1}$ events J_ψ. Thus, $|\mathcal{J}_\varphi| \leqslant (c^{\ell+1})^2$.

We now apply the symmetric version of the Local Lemma. To that aim, we need to have that $(c^{\ell+1})^2 e^{-b/100} < 1/4$, which clearly holds since ℓ is fixed and c is sufficiently large. Thus, by Lovász Local Lemma, $\mathbb{P}\big[\bigcap_{\forall \varphi} \overline{J_\varphi}\big] > 0$. Thus, there exists pairings of the edges of the cones Γ such that no event J_φ occurs for every cone φ.

Let $\mathcal{P}$ be the ℓ-collection obtained from Γ by removing all bad walks and short walks. Let $R := G \backslash U_\mathcal{P}$. We can assume that L is sufficiently large so that $\ell^4 b < \varepsilon(1-\varepsilon)c/2$. Then:

(i) In every cone φ, there are no more than εc bad and short walks via it, so there are at least $(1-\varepsilon)c$ paths in $\mathcal{P}$ via it. Hence, using the fact that G is nearly equitably colored and by considering the special 1-half and ℓ-half-cones, we obtain that for every vertex v, there are at least $\frac{1-\varepsilon}{2\ell} d_G(v)$ paths in $\mathcal{P}$ starting at v, and at least $\frac{1-\varepsilon}{2\ell} d_G(v)$ paths in $\mathcal{P}$ ending at v. Hence, $d_\mathcal{P}(v) \geqslant \frac{1-\varepsilon}{\ell} d_G(v)$. The nearly equitable ℓ-edge-coloring implies immediately that $d_\mathcal{P}(v) \leqslant \frac{1+\varepsilon}{\ell} d_G(v)$.

(ii) For every pair of vertices u, v, $u \neq v$, among all walks via a cone of u, the ratio of walks going through v is less than $\ell^2 b/c < \varepsilon/2\ell$. Hence, among all walks via u, the ratio of walks going through v is less than $\varepsilon/2\ell$. Thus

$$\frac{|\{P \in \mathcal{P} : u, v \in P\}|}{d_G(u)} \leqslant \varepsilon/2\ell,$$

and, hence, $\mathrm{conf}(u) \leqslant \varepsilon$.

(iii) In every cone, there are no more than $\ell^3 b$ bad and short walks via it, so the proportion of bad walks is at most $\ell^3 b/c < \varepsilon(1-\varepsilon)/2\ell$. Hence, among all walks via a vertex v, the ratio of bad and short walks is less than $\varepsilon(1-\varepsilon)/2\ell$. Thus $d_R(v) < \varepsilon(1-\varepsilon)d_G(v)/2\ell$, implying $d_R(v) \leqslant \varepsilon d_{\mathcal{P}}(v)$.

This completes the proof of Theorem 3.9. □

In the sequel, given two collections $\mathcal{P}_1$ and $\mathcal{P}_2$ over the same graph, we will need to grow paths of, say, $\mathcal{P}_1$ using the paths from $\mathcal{P}_2$. This will essentially be achieved by considering every path P of $\mathcal{P}_1$, incident to, say, a vertex v, then considering a path P' incident to v in $\mathcal{P}_2$, and just concatenating P and P'. So that the concatenation can be performed this way for every path of $\mathcal{P}_1$, we just need $\mathcal{P}_2$ to have enough paths, and to make sure to evenly use these paths. The latter requirement can be ensured by just orienting $\mathcal{P}_2$ in a balanced way, that is so that $|d^+(v) - d^-(v)| \leqslant 1$ for every vertex v, and choosing, as P', a path out-going from v. All such out-going paths are called *private paths* of v throughout the upcoming proofs.

The collection $\mathcal{P}$ we get from G after applying Theorem 3.9 hence satisfies $\frac{1-\varepsilon}{\ell} d_G(v) \leqslant d_{\mathcal{P}}(v) \leqslant \frac{1+\varepsilon}{\ell} d_G(v)$ for every vertex v. If we preserve the orientation of the edges of H as in the proof, and denote by $d^+_{\mathcal{P}}(v)$ the number of paths starting from v in $\mathcal{P}$, we get

$$\frac{1-\varepsilon}{2\ell} d_G(v) \leqslant d^+_{\mathcal{P}}(v) \leqslant \frac{1+\varepsilon}{2\ell} d_G(v)$$

for every vertex v. These $d^+_{\mathcal{P}}(v)$ paths out-going from v will hence be regarded as its private paths in what follows.

THEOREM 3.10

Let ℓ be a positive integer, and ε' be a sufficiently small positive real number depending on ℓ. There exists L such that, for every graph G with minimum degree at least L, there is an $(\ell, \ell+1)$-collection $\mathcal{P}$ decomposing G with

- $\mathrm{conf}(\mathcal{P}) \leqslant 1/4(\ell+10)$, *and*
- $\frac{1-\varepsilon'}{\ell} d_G(v) \leqslant d_{\mathcal{P}}(v) \leqslant \frac{1+\varepsilon'}{\ell} d_G(v)$ *for every vertex v.*

Proof. Let $\varepsilon' > 0$ be sufficiently small, and set $\varepsilon := \varepsilon'/10\ell$. Let G_1 be a $1/9\ell$-fraction of G obtained by Proposition 3.4, and $G_2 := G \backslash G_1$. By applying Theorem 3.9 on G_1 and G_2 with ε, we get two ℓ-collections $\mathcal{P}_1$ and $\mathcal{P}_2$ and two remainders R_1 and R_2 satisfying all properties from the statement of Theorem 3.9. For convenience, we will keep the orientation of the edges of $\mathcal{P}_1$ and $\mathcal{P}_2$ given by Theorem 3.9. Note that

$$\tfrac{1-\varepsilon}{\ell} \cdot \left(\tfrac{d_G(v)}{9\ell} - 2\right) \leqslant d_{\mathcal{P}_1}(v) \leqslant \tfrac{1+\varepsilon}{\ell} \cdot \left(\tfrac{d_G(v)}{9\ell} + 2\right)$$

and

$$\tfrac{1-\varepsilon}{\ell} \cdot \left(\tfrac{(9\ell-1)d_G(v)}{9\ell} - 2\right) \leqslant d_{\mathcal{P}_2}(v) \leqslant \tfrac{1+\varepsilon}{\ell} \cdot \left(\tfrac{(9\ell-1)d_G(v)}{9\ell} + 2\right).$$

Now, we have $\frac{1-\varepsilon}{(1+\varepsilon)(9\ell-1)} d_{\mathcal{P}_2}(v) - 10 \leqslant d_{\mathcal{P}_1}(v) \leqslant \frac{1+\varepsilon}{(1-\varepsilon)(9\ell-1)} d_{\mathcal{P}_2}(v) + 10$ for all vertices v. Let $R := R_1 \cup R_2$. Then for every vertex v, we have

$$d_R(v) = d_{R_1}(v) + d_{R_2}(v) \leqslant \varepsilon d_{\mathcal{P}_1}(v) + \varepsilon d_{\mathcal{P}_2}(v) \leqslant 10\ell\varepsilon d_{\mathcal{P}_1}(v).$$

Arbitrarily orient the edges of R. In our construction, every step consists in extending an arc vu of R using a private (i.e., out-going) ℓ-path starting at v in $\mathcal{P}_1$ that does not contain u – thus forming an $(\ell+1)$-path. Since the conflict ratio of $\mathcal{P}_1$ satisfies $\mathrm{conf}(\mathcal{P}_1) \leqslant \varepsilon$, at most $\varepsilon d_{\mathcal{P}_1}(v)$ paths in $\mathcal{P}_1$ with v as endpoint contain u. Note that the number of directed ℓ-paths in $\mathcal{P}_1$ starting at v is $d^+_{\mathcal{P}_1}(v) \geqslant \frac{1}{2} \cdot \frac{(1-\varepsilon)d_{G_1}(v)}{\ell}$. Thus, $d^+_{\mathcal{P}_1}(v) - d_R(v) > \varepsilon d_{\mathcal{P}_1}(v)$ since L can be chosen sufficiently large. Hence, all the $d_R(v)$ edges can be used to form $(\ell+1)$-paths.

We call $\mathcal{P}'_1$ the resulting $(\ell, \ell+1)$-collection obtained by concatenating paths from $\mathcal{P}_1$ and paths from R. Since $d_R(v) \leqslant 10\ell\varepsilon d_{\mathcal{P}_1}(v)$ for every v, the degree of v in $\mathcal{P}'_1$ is as

$$d_{\mathcal{P}_1}(v) - 10\ell\varepsilon d_{\mathcal{P}_1}(v) \leqslant d_{\mathcal{P}'_1}(v) \leqslant d_{\mathcal{P}_1}(v) + 10\ell\varepsilon d_{\mathcal{P}_1}(v).$$

Let $\mathcal{P} := \mathcal{P}'_1 \cup \mathcal{P}_2$. Then $\mathcal{P}$ is an $(\ell, \ell+1)$-collection decomposing G, in which we have $d_{\mathcal{P}}(v) = d_{\mathcal{P}'_1}(v) + d_{\mathcal{P}_2}(v)$ for all vertices v. Thus,

$$d_{\mathcal{P}_1}(v) - 10\ell\varepsilon d_{\mathcal{P}_1}(v) + d_{\mathcal{P}_2}(v) \leqslant d_{\mathcal{P}}(v) \leqslant d_{\mathcal{P}_1}(v) + 10\ell\varepsilon d_{\mathcal{P}_1}(v) + d_{\mathcal{P}_2}(v).$$

Thus,

$$\tfrac{1-\varepsilon}{\ell} d_G(v) - 10\ell\varepsilon d_{\mathcal{P}_1}(v) + 1 \leqslant d_{\mathcal{P}}(v) \leqslant \tfrac{1+\varepsilon}{\ell} d_G(v) + 10\ell\varepsilon d_{\mathcal{P}_1}(v) + 1.$$

Since $\varepsilon' = 10\ell\varepsilon$, we obtain that

$$\tfrac{1-\varepsilon'}{\ell} d_G(v) \leqslant d_{\mathcal{P}}(v) \leqslant \tfrac{1+\varepsilon'}{\ell} d_G(v).$$

Observe also that $d_{\mathcal{P}'_1}(v)/d_{\mathcal{P}_2}(v) \leqslant 1/6\ell$. Thus,

$$\mathrm{conf}(\mathcal{P}) \leqslant \mathrm{conf}(\mathcal{P}_2) + \mathrm{conf}(\mathcal{P}'_1)/6\ell \leqslant \varepsilon + 1/6\ell < 1/4(\ell+10),$$

as required. □

3.3 Tree-like collections

This part is the combinatorial core behind the proof of Theorem 3.2. We call $\mathcal{P}$ a *tree-like collection* if $H_{\mathcal{P}}$ is a tree (even if the paths of $\mathcal{P}$ pairwise intersect). We also say that $\mathcal{P}$ spans G if $H_{\mathcal{P}}$ spans G. We need here to show the existence of particular tree-like collections, namely $(\ell, 2\ell)$-tree-like collections, under mild connectivity and minimum degree requirements. These $(\ell, 2\ell)$-tree-like collections will play a crucial role to insure that some collection has all of its vertices being of even degree. However, directly getting an $(\ell, 2\ell)$-tree-like collection seems a bit challenging, and we will follow a long way for this, starting with a $(1, 2)$-tree-like collection and making its paths grow.

We start with the following lemma which is the key for the drop of the large edge-connectivity requirement.

LEMMA 3.11

> *Every 2-edge-connected multigraph G has a spanning $(1, 2)$-tree-like collection $\mathcal{T}$ such that $H_{\mathcal{T}}$ is subcubic.*

In this proof and the proof of following lemmas, we use a notion called *structured-tree*, which is a special digraph satisfying some specific properties. At first, a structured-tree is not a collection but will ultimately evolves to a tree-like collection.

Proof. Let G be connected and bridgeless. Let us call a *structured-tree* $\mathcal{T}$ a strongly-connected digraph whose vertices are subsets X_i of $V(G)$ satisfying the following properties:

- The X_i's form a partition of $V(G)$.
- The arcs of $\mathcal{T}$ are of two types: the *forward arcs* forming a rooted out-arborescence A, and the *backward arcs*, always directed from a vertex to one of its ancestors in A.
- Every arc X_iX_j corresponds to some edge $x_ix_j \in E(G)$ such that $x_i \in X_i$ and $x_j \in X_j$.
- There is at most one backward arc leaving each vertex X_i (unless $\mathcal{T}$ is rooted at X_i).
- Internal vertices of A are singletons.
- Every leaf X_i of A is spanned by a $(1, 2)$-tree-like collection $\mathcal{T}_i$ on G with maximum degree 3.

- The (unique) forward and backward arcs incident to a leaf X_i have endpoints in $\mathcal{T}_i$ with degree at most 2, and if these endpoints coincide, the degree is at most 1 in $\mathcal{T}_i$. In other words, adding the arcs as edges of $\mathcal{T}_i$ preserves maximum degree 3.
- Every edge of G is involved in at most one arc of $\mathcal{T}$ and one path of $\mathcal{T}_i$. In other words, the edges of G involved in $\mathcal{T}$ and the $\mathcal{T}_i$'s are distinct.

Let G be connected and bridgeless. Let us call $\mathcal{T}$ a strongly-connected digraph whose vertices are subsets X_i of $V(G)$ satisfying the following properties:

- The X_i's form a partition of $V(G)$.
- The arcs of $\mathcal{T}$ are of two types: the *forward arcs* forming a rooted out-arborescence A, and the *backward arcs*, always directed from a vertex to one of its ancestors in A.
- Every arc X_iX_j corresponds to some edge $x_ix_j \in E(G)$ such that $x_i \in X_i$ and $x_j \in X_j$.
- There is at most one backward arc leaving each vertex X_i (unless $\mathcal{T}$ is rooted at X_i).
- Internal vertices of A are singletons.
- Every leaf X_i of A is spanned by a $(1,2)$-tree-like collection $\mathcal{T}_i$ on G with maximum degree 3.
- The (unique) forward and backward arcs incident to a leaf X_i have endpoints in $\mathcal{T}_i$ with degree at most 2, and if these endpoints coincide, the degree is at most 1 in $\mathcal{T}_i$. In other words, adding the arcs as edges of $\mathcal{T}_i$ preserves maximum degree 3.
- Every edge of G is involved in at most one arc of $\mathcal{T}$ and one path of $\mathcal{T}_i$. In other words, the edges of G involved in $\mathcal{T}$ and the $\mathcal{T}_i$'s are distinct.

We first show that G has such a structured-tree $\mathcal{T}$, using a classical algorithm to find a strongly-connected orientation of a bridgeless graph. Fix a vertex x and compute a Depth-First-Search tree A from x. Orient the edges of A from x to form the forward arcs. By the DFS property, every edge of G not in A joins vertices which are parents. Orient these edges from the descendent to the ancestor: these are our backward arcs. Since we need to keep at most one backward arc issued from every vertex, we only keep the arc going to the lowest ancestor. Note that we obtain a structured-tree $\mathcal{T}$, where each X_i is a singleton vertex in G and every leaf $\mathcal{T}_i$ is a trivial $(1,2)$-tree-like collection on one vertex.

We now prove that every structured-tree $\mathcal{T}$ with at least two vertices on G can be reduced to one with less vertices. This will imply that $\mathcal{T}$ can be reduced to a single vertex $X_i = V(G)$, hence providing the subcubic spanning $(1,2)$-tree-like collection $\mathcal{T}_i$.

We start by deleting the backward arcs of $\mathcal{T}$ which are not needed for strong connectivity. Then we consider an internal vertex $X_j = \{x_j\}$ of A with maximal height. Let $X_1, X_2, ..., X_r$ be the (leaf) children of X_j. Each forward arc X_jX_i corresponds to an edge x_jx_i, where $x_j \in X_j$ and $x_i \in X_i$. Each of these leaves X_i is the origin of a backward arc X_iX_i' which we write y_ix_i', where $y_i \in X_i$ and $x_i' \in X_i'$. We assume that our enumeration satisfies that X_{i+1}' is always an ancestor of X_i' (possibly equal to X_i'). We now discuss the different reductions, in which the conditions of structured-trees are easily checked to be preserved.

- If X_j has only one child X_1 and is not the origin of a backward arc, we merge X_1 and X_j into a unique leaf X_{1j} spanned by the $(1,2)$-tree-like collection $\mathcal{T}_1 \cup \{x_1x_j\}$. If X_j is the root, we are done, otherwise we let the forward arc entering X_{1j} be the one entering X_j, and the backward arc leaving X_{1j} be $X_{1j}X_1'$ (thus corresponding to the edge y_1x_1').

- If X_j has only one child and is the origin of a backward arc X_jX_j', we merge X_1 and X_j into a unique leaf X_{1j} spanned by the $(1,2)$-tree-like collection $\mathcal{T}_1 \cup \{x_1x_j\}$. The forward arc entering X_{1j} is the one entering X_j, and the backward arc leaving X_{1j} is the one of X_jX_j'.

- If X_j has at least three children, or X_j has two children and is the origin of a backward arc, observe that deleting X_1 and X_2 from $\mathcal{T}$ preserves strong connectivity. Hence we merge X_1 and X_2 into a unique leaf X_{12} spanned by the $(1,2)$-tree-like collection $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \{x_1x_jx_2\}$. The forward arc entering X_{12} is $x_1'y_1$ (hence reversing the backward arc X_1X_1'), and the backward arc leaving X_{12} is $X_{12}X_2'$ corresponding to y_2x_2'.

- The last case is when X_j has two children X_1 and X_2 and is not the origin of a backward arc. Here we merge X_1, X_2, X_j into a unique leaf X_{12j} spanned by the $(1,2)$-tree-like collection $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \{x_1x_j\} \cup \{x_2x_j\}$. If X_j is the root, we are done, otherwise we let the forward arc entering X_{12j} be the one entering X_j, and the backward arc leaving X_{12j} be $X_{12j}X_2'$ (thus corresponding to y_2x_2').

This completes the proof of Lemma 3.11. □

We now turn our $(1,2)$-tree-like collection into a $(1,k)$-tree-like collection. For this we need to feed our original 2-edge-connected graph G (in which we find the $(1,2)$-tree-like collection) with some additional graph H, edge-disjoint from G, and with large enough degree.

LEMMA 3.12

Let $G = (V, E)$ be a graph. Let $\mathcal{T}$ be a spanning $(1, k)$-tree-like collection of G, where $k \geqslant 2$. Let H be a graph on V, edge-disjoint from G, with the property that $d_H(v) \geqslant 2(d_T(v) + 2k)$ for every vertex v of G. Then there is a $(1, k+1)$-tree-like collection $\mathcal{T}'$ spanning $G \cup H$.

Proof. Start by arbitrarily orienting the edges of H in a balanced way so that every vertex v of H has outdegree at least $d_T(v) + 2k$. Every vertex is hence provided with a set of private edges in H, namely, its out-going arcs. We will use these private edges to transform k-paths of $\mathcal{T}$ into $(k+1)$-paths.

In this proof, a *structured-tree* $\mathcal{T}'$ on G is a rooted $(1, k)$-tree-like collection whose vertices are subsets X_i partitioning $V(G)$ and satisfying the following properties:

- If X_iX_j is an edge in $\mathcal{T}'$, then there exists a corresponding 1-path or k-path $x_ix_j \in E(T)$, where $x_i \in X_i$ and $x_j \in X_j$.
- If X_j has children $X_1, ..., X_r$ in $\mathcal{T}'$ then there is a unique $x_j \in X_j$ such that $x_1x_j, ..., x_rx_j$ are the corresponding paths in $E(T)$. We call x_j the *center* of X_j.
- Every vertex X_i of $\mathcal{T}'$ is spanned by a $(1, k+1)$-tree-like collection $\mathcal{T}_i'$.

Initially, let $\mathcal{T}'$ be the structured-tree $\mathcal{T}$, where each X_i is a singleton element $\{x_i\}$ in $V(T)$. Note that all the vertices of $\mathcal{T}'$ are trivial $(1, k+1)$-tree-like collections. Our goal is to iteratively reduce $\mathcal{T}'$ to a structured-tree consisting of one single vertex X_i, hence providing a spanning $(1, k+1)$-tree-like collection $\mathcal{T}_i'$. We will always make sure that at any iteration every center x_j has at least $r + 2k$ private edges, where r is the number of children of X_j, hence guaranteeing the repetition of the process. Let us now show that $\mathcal{T}'$ can be reduced to a structured-tree with less vertices (unless $\mathcal{T}'$ is a single vertex).

We consider an internal vertex X_j of $\mathcal{T}'$ with maximal height. Let $X_1, ..., X_r$ be the (leaf) children of X_j corresponding to paths $x_1x_j, ..., x_rx_j$, where x_j is the center of X_j. If one of these paths, say x_1x_j, is an edge, we simply create a new vertex X_{1j} by concatenating X_1 and X_j and letting $\mathcal{T}_{1j}' = T_1' \cup T_j' \cup \{x_1x_j\}$. So we can assume that every x_ix_j-path has length k.

Consider X_j and one of its children, say X_1. Let y be a private neighbor of x_j which is not a vertex of the path x_jx_1. Such a y exists since x_j has at least $2k + r$ private neighbors. We distinguish two cases, in which the conditions of structured-trees are easily checked to be preserved:

- We first consider the case where y is in some X_i, $X_i \neq X_1$. Call P the $(k+1)$-path obtained by concatenating the k-path x_1x_j with the edge x_jy. We here add X_1 to

the set X_i to form the set X_{1i} which is spanned by $\mathcal{T}'_{1i} = T'_1 \cup T'_i \cup \{P\}$. Here x_j loses one private edge, but X_j has one less child.

- The second case is when $y \in X_1$. We here add X_1 to the set X_j to form the set X_{1j} which is spanned by the $(1, k+1)$-tree-like collection $\mathcal{T}'_{1j} = T'_1 \cup T'_j \cup \{x_j y\}$. Here x_j loses one private edge, but X_j has one less child.

This completes the proof of Lemma 3.12. □

The next result follows from Lemma 3.11 and repeated applications of Lemma 3.12:

COROLLARY 3.13

For every ℓ, there exists L such that if $G = (V, E)$ is a 2-edge-connected graph and H is another graph on V, edge-disjoint from G, with minimum degree at least L, then one can form a spanning $(1, \ell+1)$-tree-like collection $\mathcal{T}$ where $d_T(v) \leqslant d_H(v)$ for every vertex v.

Proof. We first apply Lemma 3.11 to get a subcubic $(1,2)$-tree-like collection $\mathcal{T}_0$ from G. Fix a sufficiently small $\varepsilon_1 > 0$. We choose a sequence of edge-disjoint subgraphs $H_1, ..., H_{\ell-1}$ of H, where each H_i is an ε_i-fraction of H, where $\varepsilon_{i+1} = 4\varepsilon_i$ for all i. Free to choose L large enough as a function of ε_1, we can clearly obtain the desired subgraphs $H_1, ..., H_{\ell-1}$ by repeatedly applying Proposition 3.4. Since L is sufficiently large, for each vertex v, we have that $d_{H_1}(v) \geqslant \varepsilon_1 L - 10\ell^\ell > 2d_{T_0}(v) + 4\ell$. Thus, by Lemma 3.12, we can use H_1 to extend $\mathcal{T}_0$ into a $(1,3)$-tree-like collection $\mathcal{T}_1$. Note that $d_{T_1}(v) \leqslant d_{T_0}(v) + d_{H_1}(v)$. Now we have that $d_{H_2}(v) \geqslant 3.5 d_{H_1}(v) > 2d_{T_1}(v) + 4\ell$, and thus, we can again use H_2 as an additional graph to extend $\mathcal{T}_1$ into a $(1,4)$-tree-like collection $\mathcal{T}_2$ with $d_{T_2}(v) \leqslant d_{T_0}(v) + d_{H_1}(v) + d_{H_2}(v)$. We iterate this process to form our $(1, \ell+1)$-tree-like collection $\mathcal{T}$. Note that

$$d_{\mathcal{T}}(v) \leqslant d_{T_0}(v) + \sum_{i=1}^{\ell-1} d_{H_i}(v) < L \leqslant d_H(v),$$

where the second to last inequality follows from the fact that we can choose ε_1 to be arbitrarily small. □

Our ultimate goal now is to find tree-like collections where the lengths of the paths are a multiple of some fixed value ℓ. One way to do so is to transform some $(1, \ell+1)$-tree-like collections into $(\ell, 2\ell)$-tree-like collections. Note that if ℓ is even, and our graphs G and H are bipartite with the same bipartition, then there is no spanning $(\ell, 2\ell)$-tree-like collection since an even path always connects a partite set with itself. The next result asserts that we can nevertheless connect each partite set separately.

LEMMA 3.14

For every even integer ℓ, there exists L such that if $G = (V, E)$ is a 2-edge-connected bipartite graph with vertex partition (A, B) and H is another bipartite graph on V with vertex partition (A, B), edge-disjoint from G, and with minimum degree at least L, then one can form an $(\ell, 2\ell)$-tree-like collection $\mathcal{T}$ spanning A where $d_{\mathcal{T}}(v) \leqslant d_H(v)$ for every vertex v.

Proof. We first use a small ε-fraction of H (and still call H the graph minus this fraction for convenience) in order to apply Corollary 3.13. We can then obtain a spanning $(1, \ell+1)$-tree-like collection $\mathcal{T}$ where $d_T(v) \leqslant \varepsilon d_H(v)$ for all vertices v. Note that $\varepsilon > 0$ can be taken arbitrarily small since we can take L so that εL is sufficiently large to apply Corollary 3.13. We now apply Theorem 3.9 on H to find an $(\ell-1)$-collection H' (while preserving the balanced orientation given by the proof) on H with $\mathrm{conf}(H') \leqslant \varepsilon$ and

$$\frac{1-\varepsilon}{\ell-1} d_H(v) \leqslant d_{H'}(v) \leqslant \frac{1+\varepsilon}{\ell-1} d_H(v)$$

for all vertices v.

In our construction, every step consists in extending a path P of $\mathcal{T}$ starting at some vertex v using a private (i.e., out-going) $(\ell-1)$-path from H'. This will form either an ℓ-path or a 2ℓ-path. According to the conflict ratio assumption and the fact that ε can be chosen to be sufficiently small, every such P is conflicting with at most $|P|\varepsilon d_{H'}(v) < d_{H'}(v)/8$ private paths of v, which is $1/4$ total number of private paths of v. In our upcoming process, the total number of private paths of v we will use is at most $d_T(v) \leqslant \varepsilon d_H(v)$, thus at most $1/4$ of the total number of private paths of v since $d^+_{H'}(v) \geqslant \frac{1-\varepsilon}{2\ell} d_H(v)$. Hence, even if we have already used $1/4$ of the private paths of v, and we need a private path of v which is non-conflicting with two paths of $\mathcal{T}$ incident to v, we can still find one. Thus, in the upcoming arguments, we always assume that a private path is available whenever we need one.

We now turn to the construction of the $(\ell, 2\ell)$-tree-like collection $\mathcal{T}'$ spanning A. A *structured-tree* $\mathcal{T}'$ on G is a rooted tree in which the vertices are disjoint subsets X_i whose union covers a subset of $V(G)$ containing A with the following properties:

- If X_iX_j is an edge in $\mathcal{T}'$, then there exists a corresponding 1-path or $(\ell+1)$-path $x_ix_j \in E(T)$, where $x_i \in X_i$ and $x_j \in X_j$.
- If X_j has children $X_1, ..., X_r$ then there is a unique $x_j \in X_j$ such that $x_1x_j, ..., x_rx_j$ are the corresponding paths in $\mathcal{T}$. We call x_j the *center* of X_j.
- Every vertex X_i containing an element of B is a singleton, i.e., $X_i = \{x_i\}$.
- Every vertex X_i of $\mathcal{T}'$ is spanned by an $(\ell, 2\ell)$-tree-like collection $\mathcal{T}'_i$.

We again start with $\mathcal{T}'$ equal to $\mathcal{T}$ in the sense that all X_i's are singletons, and all $\mathcal{T}_i'$'s are trivial $(\ell, 2\ell)$-tree-like collections. We root $\mathcal{T}'$ at some arbitrary vertex of A. Again our goal is to show that we can reduce $\mathcal{T}'$ until it is reduced to its root, which will therefore be equal to the set A, covered by an $(\ell, 2\ell)$-tree-like collection. Note that since ℓ is even, we always have that an edge X_iX_j of $\mathcal{T}'$ connects a vertex of B and a subset of A.

Observe first that if $\mathcal{T}'$ has a leaf in B, we can simply delete it and keep our properties. We can then assume that all leaves are subsets of A. We consider an internal vertex X_j of $\mathcal{T}'$ with maximal height. Let $X_1, X_2, ..., X_r$ be the (leaf) children of X_j corresponding to the paths $x_1x_j, ..., x_rx_j$. Note that all X_i's are subsets of A, and that $X_j = \{x_j\}$ is in B. We now discuss the different reductions, in which the conditions of structured-trees are easily checked to be preserved.

Consider X_j and one of its children, say X_1. Let X_k be the parent of X_j in $\mathcal{T}'$. Note that X_k is a subset of A. We denote by x_jx_k the path of $\mathcal{T}$ joining X_j and X_k. Let y be a private neighbor of x_j which is not a vertex of the path x_jx_k and x_jx_1. We again consider two cases:

- First assume that y is in some X_i, with $X_i \neq X_1$. We denote by P' the path obtained by concatenating the path x_1x_j with x_jy. Note that P' is an ℓ-path or a 2ℓ-path. We add X_1 to the set X_i to form the set X_{1i} which is spanned by $\mathcal{T}_{1i}' = T_1' \cup T_i' \cup \{P'\}$. Note that x_j loses a private path, but X_j has one less child.
- Otherwise, $y \in X_1$. We add X_1 to the set X_k to form the set X_{1k} which is spanned by the $(\ell, 2\ell)$-tree-like collection $\mathcal{T}_{1k}' = T_1' \cup T_k' \cup \{P''\}$, where P'' is the concatenation of x_kx_j and x_jy (note that P'' is an ℓ-path or a 2ℓ-path). Note that x_j loses a private path, but X_j has one less child.

This completes the proof of Lemma 3.14 □

We will also need the following lemma.

LEMMA 3.15

Let ℓ be a positive integer. There exists L such that if $G_1 = (V, E)$ is a 2-edge-connected graph and $G_2 = (V, F)$ is a graph of minimum degree at least L edge-disjoint from G_1, then there is a connected $[\ell, \ell + 3]$-collection $\mathcal{P}$ decomposing $G_1 \cup G_2$ with $\text{conf}(\mathcal{P}) < \frac{1}{2(\ell+10)}$.

Proof. Start by applying Lemma 3.11 to get a spanning $(1, 2)$-tree-like collection $\mathcal{T}$ of G_1 such that $H_{\mathcal{T}}$ is subcubic, and put the non-used edges of G_1 in G_2. Still calling this graph G_2, we decompose G_2 into a $1/(5\ell)$-fraction R_1 and a $1 - 1/(5\ell)$-fraction R_2, by Proposition 3.3. Thus, by Theorem 3.10, G_2 can then be decomposed into two

$(\ell, \ell+1)$-collections $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, both having conflict ratio at most $\frac{1}{4(\ell+10)}$, and verifying

$$\frac{1-\varepsilon}{(5\ell-1)(1+\varepsilon)} d_{\mathcal{P}_2}(v) \leqslant d_{\mathcal{P}_1}(v) \leqslant \frac{1+\varepsilon}{(5\ell-1)(1-\varepsilon)} d_{\mathcal{P}_2}(v)$$

for all vertices v, and any ε.

In our construction, every step consists in extending a path P of $\mathcal{T}$ starting at v using a private $(\geqslant \ell)$-path starting at v in $\mathcal{P}_1$ (where we recall that the private paths at any vertex are its out-going paths in a balanced orientation of $\mathcal{P}_1$). This will form a $(\geqslant \ell)$-path. By the assumption on the conflict ratio, every P is conflicting with at most, say, half of the private paths of v. Because $H_{\mathcal{T}}$ is subcubic, the total number of private paths of v we will need is at most 6. Since L can be chosen so that $\frac{L}{5\ell} \cdot \frac{1}{2\ell} \cdot \left(1 - \frac{1}{4(\ell+10)}\right)$ is arbitrarily large, we can hence assume we have enough private paths for the whole process.

We now turn to the construction of the spanning $(\geqslant \ell)$-tree-like collection $\mathcal{T}'$ from $\mathcal{T}$ and $\mathcal{P}_1$. We start with a *structured-tree* $\mathcal{T}'$, which is a rooted tree in which the vertices are disjoint subsets X_i partitioning V with the following properties:

- If X_iX_j is an edge in $\mathcal{T}'$, then there exists a corresponding 1-path or 2-path $x_ix_j \in E(T)$, where $x_i \in X_i$ and $x_j \in X_j$.
- Every vertex X_i of $\mathcal{T}'$ is spanned by a $(\geqslant \ell)$-tree-like collection $\mathcal{T}_i'$.

We again start with $\mathcal{T}'$ being equal to $\mathcal{T}$ in the sense that all X_i's are singletons, and all $\mathcal{T}_i'$'s are trivial $(\geqslant \ell)$-tree-like collections. We root $\mathcal{T}'$ at some arbitrary vertex. Again our goal is to show that we can reduce $\mathcal{T}'$ until it is reduced to its root, which will therefore be a spanning $(\geqslant \ell)$-tree-like collection.

We consider a leaf X_1 of $\mathcal{T}'$ with direct ancestor X_j. Then there exists a path x_1x_j of $\mathcal{T}'$ having length 1 or 2. We pick a private path $x_jy \in H_1$ not conflicting with the path x_1x_j. Assume $y \in X_k$. If $X_k \neq X_1$, then we denote by P the path obtained by concatenating x_1x_j and x_jy. Then we add X_1 to X_k to form the set X_{1k} being spanned by $\mathcal{T}_{1k}' = \mathcal{T}_1' \cup \mathcal{T}_k' \cup \{P\}$. If $X_k = X_1$, then we add X_1 to X_j to form the set X_{1j} being spanned by $\mathcal{T}_{1j}' = \mathcal{T}_1' \cup \mathcal{T}_j' \cup \{x_jy\}$. We choose a private path x_jz in $\mathcal{P}_1$ not conflicting with x_1x_j, and concatenate these two paths to get a path x_1z that we put back into $\mathcal{P}_1$.

Once the procedure above is finished, we end up with a spanning $(\geqslant \ell)$-tree-like collection $\mathcal{T}'$ and an $(\ell, \ell+1)$-collection $\mathcal{P}_1'$, where $\mathcal{P}_1'$ is the collection remaining from $\mathcal{P}_1$ after we have used some of its paths to obtain $\mathcal{T}'$. Let $\mathcal{P} := \mathcal{T}' \cup \mathcal{P}_1' \cup \mathcal{P}_2$. Then $\mathcal{P}$ covers all edges of G. Note also that $\mathcal{P}$ is an $[\ell, \ell+3]$-collection. Since $d_{\mathcal{T}' \cup \mathcal{P}_1'}(v) \leqslant d_{\mathcal{P}_1}(v) + 3$ for every vertex v and we can choose ϵ to be sufficiently small, we have $d_{\mathcal{T}' \cup \mathcal{P}_1'}(v) \leqslant d_{\mathcal{P}_2}(v)/4(\ell+10)$ for every vertex v. Thus,

$$\mathrm{conf}(\mathcal{P}) \leqslant \mathrm{conf}(\mathcal{P}_2) + \frac{\mathrm{conf}(\mathcal{T}' \cup \mathcal{P}_1')}{4(\ell+10)} < \frac{1}{4(\ell+10)} + \frac{1}{4(\ell+10)} \leqslant \frac{1}{2(\ell+10)},$$

which concludes the proof. □

3.4 Path-decompositions

Before completing the proof of Theorem 3.2, we need some notions about Eulerian graphs, which we will study in details in Chapter 4. We call $\mathcal{P}$ an *Eulerian collection* if $H_{\mathcal{P}}$ is Eulerian. From an Eulerian tour in $H_{\mathcal{P}}$, we naturally get a corresponding *Eulerian tour* in $\mathcal{P}$. Such a tour is *non-conflicting* if every two of its consecutive paths are non-conflicting. It was proved, under the following different terminology, by Jackson (cf. [Jac93], Theorem 6.3) that Eulerian collections with somewhat low conflicts have non-conflicting Eulerian tours.

The original result of Jackson is as follows. For a vertex v, let E_v be the set of edges incident to v. A *generalised transition system* S for a graph G is a set of functions $\{S_v\}_{v\in V(G)}$ such that $S_v : E_v \to 2^{E_v}$ and whenever $e_1 \in S_v(e_2)$, we have that $e_2 \in S_v(e_1)$. We say that an Eulerian tour $\mathcal{E}$ is *compatible* with S if for all $v \in V(G)$, whenever $e_1 \in S_v(e_2)$ it follows that e_1 and e_2 are not consecutive edges in $\mathcal{E}$.

THEOREM 3.16 Jackson [Jac93]

Let S be a generalised transition system for an Eulerian graph G. Suppose that for each vertex $v \in V(G)$ and $e \in E_v$, we have

(i) $|S_v(e)| \leqslant \frac{1}{2}d(v) - 1$ *when* $d(v) \equiv 0 (mod\, 4)$ *or* $d(v) = 2$, *and*

(ii) $|S_v(e)| \leqslant \frac{1}{2}d(v) - 2$, *otherwise.*

Then G has an Eulerian tour compatible with S.

From Theorem 3.16, the following result is immediate.

THEOREM 3.17

Every Eulerian collection $\mathcal{P}$ with $\mathrm{conf}(\mathcal{P}) \leqslant 1/2(\ell + 10)$ *and* $d_{\mathcal{P}}(v) \geqslant \ell + 10$ *for every v has a non-conflicting Eulerian tour.*

Proof. Let $P \in \mathcal{P}(v)$. The number of paths of $\mathcal{P}(v)$ conflicting with P is at most $\frac{1}{2(\ell+10)}(\ell + 3)d_{\mathcal{P}}(v)$, and so at most $\frac{1}{2}d_{\mathcal{P}}(v) - 2$ since $d_{\mathcal{P}}(v) \geqslant \ell + 10$. The result now follows from Theorem 3.16. □

We now have all ingredients to prove Theorems 3.2.

Proof of Theorem 3.2. Without loss of generality, we assume that ℓ is even (as the statement for paths of length $2k$ implies the statement for paths of length k). First of all,

we consider a maximum cut (V_1, V_2) of G, and just keep the set of edges F across the cut. We call G' the graph (V, F). Observe that G' is at least 12-edge-connected and has minimum degree at least $d_\ell/2$.

By Proposition 3.6, there is an orientation D of G' such that D is 6-arc-strong and with $d^+(v)$ and $d^-(v)$ differing by at most 1 for every vertex v. By applying Proposition 3.7 to D with some vertex z, we obtain 6 arc-disjoint out-arborescences, $T_1, ..., T_6$, rooted at z. Since each vertex v has in-degree at most 1 in T_i (z has in-degree 0), and $d^+_D(v)$ and $d^-_D(v)$ differ by at most 1, the graph $T_1 \cup ... \cup T_6$ is $1/2$-sparse in G'.

Call now $G_1 := T_1 \cup T_2$, $G_2 := T_3 \cup T_4$, $G_3 := T_5 \cup T_6$, and let R be the graph consisting of all the edges of F which are not in G_1, G_2, G_3. Observe that G_1, G_2, G_3 are connected and bridgeless. Furthermore, the graph $G_1 \cup G_2 \cup G_3$ is $1/2$-sparse in G', and hence R is $1/2$-dense in G'. In the sequel, several fractions of edges will be removed from R, but, for the sake of legibility, we will still call R the set remaining after the transfers.

We turn G_1 into an $(\ell, 2\ell)$-tree-like collection as follows: we consider a small ε-fraction R_1 of R, and apply Lemma 3.14 (with G_1 for G and R_1 for H) to form an $(\ell, 2\ell)$-tree-like collection $\mathcal{T}'$ spanning V_1 in which $d_{\mathcal{T}'}(v) \leqslant d_{R_1}(v)$ for all vertices $v \in V_1$. In other words, $\mathcal{T}'$ is ε'-sparse in R for some negligible $\varepsilon' > 0$ depending on ε. Similarly, we can obtain, from G_2, a ε'-sparse $(\ell, 2\ell)$-tree-like collection $\mathcal{T}''$ spanning V_2. We still consider (neglecting the two ε-fractions) that R is $1/2$-dense in G'. Add all edges of $E(G)\setminus F$ to R.

Now, $G = G_3 \cup U_{\mathcal{T}'} \cup U_{\mathcal{T}''} \cup R$. We claim that we can remove a collection of ℓ-paths or 2ℓ-paths from the tree-like collection $\mathcal{T}'$ spanning V_1 in a way so that we can obtain that at most one vertex of V_1 has odd degree in G. Indeed, note that the following claim is straightforward.

Given a tree T and an even subset X of $V(T)$, then there exists a set of edges $F \subseteq E(T)$ such that for each vertex x, $d_F(x)$ is odd if and only if $x \in X$ (one way to see this is to note that the characteristic vector of X is in the span of the incidence matrix of T).

In particular, denoting by X_1 the set of all odd-degree vertices of $G_3 \cup U''_{\mathcal{T}} \cup R$ inside V_1 (and possibly removing one vertex of X_1 to make X_1 of even size) we can find a sub-collection $\mathcal{F}'$ of $\mathcal{T}'$ such that $d_{\mathcal{F}'}(v)$ is odd if and only if $v \in X_1$. In other words, removing the ℓ- or 2ℓ-paths of $\mathcal{T}'$ corresponding to $\mathcal{F}'$ leaves G with every vertex of V_1 (except possibly one) having even degree. Similarly, we remove paths of the tree-like collection $\mathcal{T}''$ spanning V_2 so that at most one vertex of V_2 has odd degree.

We still call G the remaining graph after the procedure, and we add the remaining edges of $U_{\mathcal{T}'} \cup U_{\mathcal{T}''}$ to R. Then $G = G_3 \cup R$. Note that G_3 is 2-edge-connected, and R is $1/4$-dense in G. By applying Lemma 3.15 (with G_3 for G_1 and R for G_2), G contains a connected $[\ell, \ell+3]$-collection $\mathcal{P}$ with $\text{conf}(\mathcal{P}) < 1/2(\ell+10)$. Note that $d_G(v) - d_{\mathcal{P}}(v)$ is even for every vertex v – so the degree of every vertex in $\mathcal{P}$ is even, except (possibly) for two vertices $v_1 \in V_1$ and $v_2 \in V_2$. In this case, we add a dummy ℓ-path from v_1

to v_2 in $\mathcal{P}$ to make H Eulerian. By Theorem 3.17, we get that $\mathcal{P}$ has a non-conflicting Eulerian tour from which we can deduce the desired decomposition. □

Chapter 4

Decomposing Eulerian Graphs

4.1 Häggkvist and Kriesell's conjecture

In this chapter, we will investigate pattern-decompositions of large Eulerian graphs. Another equivalent definition of an Eulerian graph is that it is connected and the degree of all vertices are even, which is relatively cheap to achieved just by arbitrarily pairing odd-degree vertices and add an edge between each pair. Thank to the even-degree property, Eulerian graphs offers a significant advantage for both flows in graphs and path-decomposition, since the main task in these problems is pairing incident edges. That advantage is verified by the fact that *a connected graph has a* 2*-flow if and only if it is Eulerian*. Besides, the proof of Theorem 3.2 in Section 3.4 also employs some arguments on Eulerian graphs. In fact, it is not very hard to see that if the graph is Eulerian, then we can reduce the edge-connectivity to 4.

THEOREM 4.1

> *For every positive integer ℓ, there is an integer d'_ℓ such that every* 4*-edge-connected Eulerian graph G with minimum degree at least d'_ℓ and size divisible by ℓ admits a P_ℓ-decomposition.*

The proof of Theorem 4.1 will be presented in Section 4.3. It is natural to ask whether we can remove the 4-edge-connected condition, i.e, whether *every Eulerian graph with high minimum degree and size divisible by ℓ admits a P_ℓ-decomposition.* It turns

The content of this chapter is covered in paper *Locally self-avoiding Eulerian tours,* Tien-Nam Le, *Journal of Combinatorial Theory, Series B*, to appear, and in a part of paper *Edge-partitioning a graph into paths: beyond the barát- thomassen conjecture,* Julien Bensmail, Ararat Harutyunyan, Tien-Nam Le, and Stéphan Thomassé, *Combinatorica*, pages 1–25, 2018.

out that Häggkvist ([Hag89], Problem 3.3) and Kriesell [Kri11] conjectured a more general statement regarding self-avoiding Eulerian tours.

Given an Eulerian tour $\mathcal{E}$ of a multigraph G, for every positive integer ℓ, a walk $e_1e_2...e_\ell$ where any e_i, e_{i+1} are consecutive edges of $\mathcal{E}$ is called a *segment* of *length* ℓ of $\mathcal{E}$. We say that $\mathcal{E}$ is *ℓ-step self-avoiding* if every segment of length at most ℓ of $\mathcal{E}$ is a path, which is equivalent to that $\mathcal{E}$ "contains" no cycle of length at most ℓ.

Häggkvist ([Hag89], Problem 3.3) and Kriesell [Kri11] independently conjectured that high minimum degree is a sufficient condition for the existence of an ℓ-step self-avoiding Eulerian tour.

CONJECTURE 4.2 [Hag89, Kri11]

> *For every positive integer ℓ, there is an integer d_ℓ such that every Eulerian graph G with minimum degree at least d_ℓ admits an ℓ-step self-avoiding Eulerian tour.*

Häggkvist also asked to identify the minimum of d_ℓ if it exists. There is in fact very little is known about self-avoiding Eulerian tour, except for the case $\ell = 3$, a.k.a. *triangle-free Eulerian tours*. Adelgren [Ade95] characterized all graphs with maximum degree at most 4 which admit a triangle-free Eulerian tour before Oksimets [Oks97] proved Conjecture 4.2 for $\ell = 3$ with a sharp bound $d_3 = 6$. The main part of this chapter is devoted to verify Conjecture 4.2.

THEOREM 4.3

> *For every positive integer ℓ, there is an integer d_ℓ such that every Eulerian graph G with minimum degree at least d_ℓ admits an ℓ-step self-avoiding Eulerian tour.*

By just cutting the tour found by Theorem 4.3 into paths of length ℓ, we immediately obtain a P_ℓ-decomposition of G, answering the question at the beginning of this chapter.

COROLLARY 4.4

> *For every integer $\ell \geqslant 2$, there is an integer d_ℓ such that every Eulerian graph with size divisible by ℓ and minimum degree at least d_ℓ can be decomposed into paths of length ℓ.*

Outline the proof of Theorem 4.3

The following theorem is the self-avoiding tour version of Theorem 4.1.

THEOREM 4.5

> *For every positive integer ℓ, there is an integer d'_ℓ such that every 4-edge-connected Eulerian graph G with minimum degree at least d'_ℓ admits an ℓ-step self-avoiding Eulerian tour.*

Clearly, Theorem 4.5 cannot be extended to multigraphs: a multigraph consisting of two vertices linked by many edges is a counterexample. However, the main tool to prove Theorem 4.3 is indeed the following weak extension of Theorem 4.5 to multigraphs. Roughly speaking, we only require the Eulerian tour to behave well on a given simple subgraph, not necessary on the whole multigraph. Please note that *multigraphs in this chapter may contain multiple edges and loops*, where each loop contributes two to the degree of the incident vertex.

THEOREM 4.6

> *For every integer ℓ, there is an integer d_ℓ such that for every 4-edge-connected Eulerian multigraph G with minimum degree at least d_ℓ and every simple subgraph G' of G, the multigraph G admits an Eulerian tour in which every segment of length at most ℓ and consisting of only edges of G' is a path.*

4.2 Merging multiple Eulerian tours

In this section, we use Theorem 4.6 as a black box to prove Theorem 4.3. Then we prove Theorem 4.6 in the next section. We start by recalling the definition of cactus graphs.

DEFINITION 4.7

> *A connected loopless multigraph is a* cactus *if every edge belongs to at most one cycle.*

The singleton graph is a cactus by convention. Clearly, if a cactus is Eulerian then every edge belongs to exactly one cycle. The following is a well-known property of cactus graphs.

PROPOSITION 4.8

> *There are at most two edge-disjoint paths between any two distinct vertices of a cactus.*

The main idea of the proof of Theorem 4.3 is as follows. We first partition the original graph G into 4-edge-connected Eulerian "induced subgraphs"; these subgraphs are structurally linked by a big cactus. We then apply Theorem 4.6 to obtain a well-behaved Eulerian tour of each subgraph, and finally connect these tours by the cactus to get an Eulerian tour of G.

Given a multigraph $G = (V, E)$, to *contract* a set of vertices $X \subset V$, we remove all edges inside X, and then merge the vertices of X to a new vertex x, and each edge incident to x corresponds to an edge incident with some $v \in X$. Note that if the sum of degrees of vertices of X is even, then the degree of x is even.

Let $G = (V, E)$ be an Eulerian multigraph and $\mathcal{X}$ be a partition of V into non-empty sets $X_1, X_2, ..., X_k$ for some positive integer k. Let $M_{\mathcal{X}}$ be the loopless multigraph obtained from G by contracting each X_i to a new vertex x_i. Clearly, the degree of each x_i of $M_{\mathcal{X}}$ is even. If $k \geqslant 2$, we have that $M_{\mathcal{X}}$ is connected since G is connected, and hence $M_{\mathcal{X}}$ is Eulerian.

Let us suppose for the moment that $M_{\mathcal{X}}$ is a cactus. Thus an edge e of $M_{\mathcal{X}}$ belongs to exactly one cycle in $M_{\mathcal{X}}$. Let e' be an edge of the same cycle and incident with e. We say that $\{e, e'\}$ is a *pair* at x_i, where x_i is some endpoint shared by e and e'. Note that every edge belongs to exactly one pair at each of its endpoints, and hence belongs to exactly two pairs in total. Since each edge e of $M_{\mathcal{X}}$ corresponds to an edge of G, we may use e to denote both interchangeably. For every pair $\{e, e'\}$ at some x_i, each edge has a unique endpoint in X_i, say u and u' respectively. We create a new *dummy edge* $f = uu'$ *associated with* the pair $\{e, e'\}$ (note that f may be a loop). For every $1 \leqslant i \leqslant k$, let F_i be the edge set of $G[X_i]$ and $\overline{F}_i$ be the set of all dummy edges on X_i, and let $G_i = (X_i, F_i \cup \overline{F}_i)$. We say that the multigraphs $G_1, ..., G_k$ are *inherited from* $\mathcal{X}$. Clearly, $d_{G_i}(v) = d_G(v)$ for every $v \in X_i$. The following lemma asserts that there is a partition such that the inherited multigraphs are 4-edge-connected and Eulerian, which are essential conditions to employ Theorem 4.6. For the sake of clarity, we do not consider edge-connectivity of multigraphs on a single vertex.

LEMMA 4.9

> *Given an Eulerian multigraph $G = (V, E)$, there exists a partition $\mathcal{X}$ of V such that $M_{\mathcal{X}}$ is a cactus, and every G_i inherited from $\mathcal{X}$ is either a single vertex with loops or a 4-edge-connected Eulerian multigraph.*

Proof. The proof is by induction on $|V|$. For the case $|V| = 2$, let $V = \{u, v\}$. If G has only two edges between u and v, then $\mathcal{X} = \{\{u\}, \{v\}\}$; otherwise, $\mathcal{X} = \{\{u, v\}\}$. The lemma holds for $|V| = 2$.

For the case $|V| > 2$, if G is 4-edge-connected, then $\mathcal{X} = \{V(G)\}$. Otherwise, G contains an edge-cut of size 2, i.e. an edge-cut consisting of two edges. Consider an edge-cut partitioning V into X_1 and V' such that $|X_1|$ is minimum among all possible edge-cuts of size 2. Let us call the two edges of the cut u_1v_1 and u_2v_2, where $u_1, u_2 \in X_1$ and $v_1, v_2 \in V'$. We create two dummy edges $f = u_1u_2$ and $f' = v_1v_2$. Let $\overline{F}_1 = \{f\}$, and $F_1 = E(G[X_1])$. Let $G_1 = (X_1, F_1 \cup \overline{F}_1)$ and $G' = (V', E(G[V']) \cup \{f'\})$. There are at least two edge-disjoint paths in G between any two distinct vertices of X_1. If both paths contains vertices of V', then the edge-cut must has size at least 4, a contradiction. Therefore there is a path in $G[X_1]$ between any two distinct vertices of

X_1. Thus if $|X_1| > 1$ then G_1 is connected, and hence is Eulerian since the degree of every vertex of G_1 is even. Similarly, G' is Eulerian.

Suppose that G_1 contains an edge-cut of size 2 partitioning X_1 into X_1' and X_1''. If u_1 and u_2 are in the same partition, say X_1', then that edge-cut is also an edge-cut of G partitioning V into X_1'' and $V' \cup X_1'$, which contradicts the minimality of $|X_1|$. If $u_1 \in X_1'$ and $u_2 \in X_1''$ then that edge-cut consists of f and another edge, say e. Then $\{e, u_1v_1\}$ is an edge-cut of G partitioning V into X_1' and $V' \cup X_1''$, a contradiction again. It follows that G_1 contains no edge-cut of size 2, and so is 4-edge-connected.

Applying induction hypothesis to the Eulerian multigraph G' gives a partition of V' into $\mathcal{X}' = \{X_2, ..., X_k\}$ such that $M_{\mathcal{X}'}$ and $G_2, ..., G_k$ inherited from $\mathcal{X}'$ satisfy Lemma 4.9. Let $x_i \in M_{\mathcal{X}'}$ corresponds to X_i for every $2 \leqslant i \leqslant k$. Set $\mathcal{X} = \mathcal{X}' \cup \{X_1\}$ and construct $M_{\mathcal{X}}$ as follows:

(a) If $v_1, v_2 \in G_i$ for some i, then $M_{\mathcal{X}}$ is obtained from $M_{\mathcal{X}'}$ by adding x_1 and two parallel edges x_1x_i, corresponding to edges u_1v_1 and u_2v_2 of G. Hence there is only one pair at x_1: $\{u_1v_1, u_2v_2\}$, and f is its associated dummy edge. There is one more pair at x_i in $M_{\mathcal{X}}$ comparing with x_i in $M_{\mathcal{X}'}$: $\{v_1u_1, v_2u_2\}$, and f' is its associated dummy edge.

(b) Otherwise, $v_1 \in G_i$ and $v_2 \in G_j$ for some $i \neq j$. There must be an edge x_ix_j in $M_{\mathcal{X}'}$ corresponding to f' in G'. We obtain $M_{\mathcal{X}}$ from $M_{\mathcal{X}'}$ by adding vertex x_1, edge x_1x_i corresponding to u_1v_1 and edge x_1x_j corresponding to u_2v_2 together with deleting the edge x_ix_j corresponding to f'. There is only one pair at x_1: $\{u_1v_1, u_2v_2\}$, and f is its associated dummy edge. The set of pairs at x_i (res. x_j) of $M_{\mathcal{X}}$ are identical to the set of pairs at x_i (res. x_j) of $M_{\mathcal{X}'}$, except that v_1u_1 (res. v_2u_2) replaces f' in some pair at x_i (res. at x_j).

The multigraphs $G_2, ..., G_k$ inherited from $\mathcal{X}$ in this construction are identical to the multigraphs $G_2, ..., G_k$ inherited from $\mathcal{X}'$. By induction hypothesis, for every $i \geqslant 2$, if G_i has more than one vertex then it is 4-edge-connected and Eulerian. Note that x_1 has degree 2, and $M_{\mathcal{X}'}$ is a cactus, then so is $M_{\mathcal{X}}$. This proves the lemma. □

Given an Eulerian tour $\mathcal{E}$ of G and a subset X of V, a segment $v_1v_2...v_r$ ($r \geqslant 3$) of $\mathcal{E}$ is an *X-boomerang* if $v_1, v_r \in X$ and $v_2, ..., v_{r-1} \notin X$. A *projection* of $\mathcal{E}$ on X is an Eulerian tour $\mathcal{E}_X$ obtained from $\mathcal{E}$ by replacing every X-boomerang, say $v_1v_2...v_r$, by a dummy edge (possibly a loop) between v_1 and v_r. If $\mathcal{E}_X$ is a projection of $\mathcal{E}$, we say $\mathcal{E}$ and $\mathcal{E}_X$ are *compatible*.

Let G be an Eulerian multigraph and $\mathcal{X}$ be a partition of G together with $M_{\mathcal{X}}$ and inherited $G_1, ..., G_k$ given by Lemma 4.9. For every i, let $\mathcal{E}_i$ be an arbitrary Eulerian tour of G_i.

CLAIM 4.10

> *There exists an Eulerian tour $\mathcal{E}$ of G compatible with all $\mathcal{E}_i$. Furthermore, for every pair $\{e, e'\}$ at some x_i, there is an X_i-boomerang of $\mathcal{E}$ starting and ending by e and e'.*

Proof. We reuse all notations in the proof of Lemma 4.9 and proceed by induction on k. The claim clearly holds for $k = 1$. For $k > 1$, recall that by the algorithm in the proof of Lemma 4.9, the Eulerian multigraph G' has $k-1$ inherited multigraphs identical to $G_2, ..., G_k$ of G. Hence, by the induction hypothesis applied on G', there exists an Eulerian tour $\mathcal{E}'$ of G' compatible with all $\mathcal{E}_i, i \geqslant 2$, and for every pair $\{e, e'\}$ at some $x_i, i \geqslant 2$, there is an X_i-boomerang of $\mathcal{E}'$ starting and ending by e and e'. Note that in cases (a) and (b) of the proof of Lemma 4.9, the only pair at x_1 is $\{u_1v_1, u_2v_2\}$ associated with f. Let W_1 be the walk obtained from $\mathcal{E}_1$ by removing f, and $\mathcal{E}$ be the Eulerian tour on G obtained from $\mathcal{E}'$ by replacing f' by the segment $v_1u_1W_1u_2v_2$. It is straightforward that, in both cases (a) and (b), the tour $\mathcal{E}$ satisfies Claim 4.10. □

Let $\{e, e'\}$ be a pair at some x_i, and W be the X_i-boomerang of $\mathcal{E}$ starting and ending by e and e'. Let $\overline{W}$ be the segment obtained from $\mathcal{E}$ by removing W.

CLAIM 4.11

> *If W visits a vertex $v \notin X_i$, then $\overline{W}$ does not visit v.*

Proof. Suppose that the claim was false. Let $v \in X_j$ for some $j \neq i$. Contracting every X_i to x_i naturally yields from W and $\overline{W}$ two edge-disjoint walks $W_{\mathcal{X}}$ and $\overline{W}_{\mathcal{X}}$ in $M_{\mathcal{X}}$, respectively. By following $W_{\mathcal{X}}$ from x_i to x_j and return to x_i, and then following $\overline{W}_{\mathcal{X}}$ to x_j, we obtain three edge-disjoint walks between x_i and x_j, contrary to Proposition 4.8. □

CLAIM 4.12

> *If G has minimum degree d, then whenever $\mathcal{E}$ leaves X_i, it takes at least d steps to return to X_i.*

Proof. The claim is equivalent to that every X_i-boomerang W has length at least d. Suppose that W visits vertex $v \notin X_i$. By Claim 4.11, W must contains all edges incident with v, and hence has length at least d. □

We are ready to prove Theorem 4.3.

Proof of Theorem 4.3. Let G be an Eulerian graph with minimum degree at least d_ℓ, the constant of Theorem 4.6. There is a partition $\mathcal{X} = \{X_1, ..., X_k\}$ of $V(G)$ together with inherited multigraphs $G_1, ..., G_k$ satisfying Lemma 4.9.

If G_i consists of only one vertex and some loops, let $\mathcal{E}_i$ be an arbitrary Eulerian tour of G_i. Otherwise, Lemma 4.9 asserts that G_i is Eulerian, 4-edge-connected, and $d_{G_i}(v) = d_G(v) \geqslant d_\ell$ for any $v \in X_i$. Also note that $G[X_i]$ is a simple subgraph of G_i. We thus get, by Theorem 4.6, an Eulerian tour $\mathcal{E}_i$ of G_i of which every segment of length at most ℓ and containing only edges of $G[X_i]$ is a path. Claim 4.10 gives an Eulerian tour $\mathcal{E}$ of G compatible with all $\mathcal{E}_i$.

The proof is completed by showing that every segment W of length at most ℓ of $\mathcal{E}$ is a path. Suppose that $W = W_1 e_1 W_2 e_2 ... e_{t-1} W_t$, where each W_s (possibly of length 0) contains only vertices of some X_{i_s}, and e_s is an edge between two distinct sets X_{i_s} and $X_{i_{s+1}}$. By Claim 4.12, whenever $\mathcal{E}$ leaves some X_{i_s}, it takes at least $d_\ell > \ell$ steps to return to X_{i_s}, while the length of W is at most ℓ. Therefore $X_{i_s} \neq X_{i_r}$ for every $s \neq r$. Because $\mathcal{E}$ is compatible with $\mathcal{E}_{i_s}$, and W_s contains only vertices of X_{i_s}, we have that W_s is a segment of $\mathcal{E}_{i_s}$. Since $W_s \subseteq G[X_{i_s}]$ and has length at most ℓ, it is a path by Theorem 4.6. This means that W is a path, and the proof is complete. □

4.3 Self-avoiding tours on Eulerian multigraphs

In this section, we will prove Theorem 4.6 to complete the proof of Theorem 4.3. We start with the proof of Theorem 4.5.

Proof of Theorem 4.5. Following the arguments in the second paragraph of Theorem 3.2, we can extract from G two trees T_1 and T_2 so that $T_1 \cup T_2$ is $1/2$-sparse in G. Let $G_1 := T_1 \cup T_2$, and $G_2 := G \backslash G_1$. Then G_1 is 2-edge-connected, and G_2 is $1/2$-dense. Applying Lemma 3.15, we can express G as a connected $[\ell, \ell+3]$-collection $\mathcal{P}$ with $\mathrm{conf}(\mathcal{P}) < 1/2(\ell+10)$. Since G is Eulerian, so is $\mathcal{P}$. Hence $\mathcal{P}$ has non-conflicting Eulerian tours according to Theorem 3.17, and these tours do not have cycles of length at most ℓ since all paths of $\mathcal{P}$ have length at least ℓ. □

Suppose that we have a decomposition $\mathcal{P}$ of an Eulerian graph G with all paths of length at least ℓ. Then just by concatenating the paths arbitrarily, we obtain a decomposition of G into several circuits since G is Eulerian. If every two consecutive paths (i.e., they are concatenated) are non-conflicting, then all circuits are ℓ-step self-avoiding. Theorem 3.10 provides a low conflicting decomposition for this purpose. In order to obtain an ℓ-step self-avoiding Eulerian tour, it is necessary that the process of concatenating returns a single circuit; this is taken care by Lemma 3.11.

The proof of Theorem 4.6 is partly similar to the proof of Theorem 4.5 but more involved. The main difficulty here is that certain arguments on simple graphs in the proof of Theorem 4.5 could not be extended to multigraphs. To overcome these difficulties, we introduce the notions of *F-paths* and *F-path-collections* as an extension of paths and path-collections in Section 3.2.

F-path-collections

Given a multigraph $G = (V, E)$ and a subgraph $G' = (V, F)$ satisfying the hypotheses of Theorem 4.6, the goal is to find an Eulerian tour $\mathcal{E}$ of G such that every segment of $\mathcal{E}$ of length at most ℓ and consisting of only edges of F is a path. To this end, we introduce a relaxation of path, called F-path, to depict the characteristics of segments of the tour. Let $G = (V, E)$ be a multigraph and F be a subset of E. A walk W in G is called an *F-path* if every subwalk of W containing only edges of F is a path. An F-path W is *covered* if all edges of W belong to F, and is *uncovered* otherwise. It is immediate that a covered F-path is a path.

An *F-collection* $\mathcal{P}$ on G is a set of edge-disjoint F-paths of G. We denote by $U_{\mathcal{P}} = (V, E')$ the multigraph where E' is the set of edges of F-paths in $\mathcal{P}$. If $U_{\mathcal{P}} = G$, then $\mathcal{P}$ is called an *F-decomposition* of G. We denote by $H_{\mathcal{P}} = (V, E'')$ the multigraph where each edge (possibly a loop) $uv \in E''$ corresponds to an F-path between u and v in $\mathcal{P}$. The *degree* of a vertex v in $\mathcal{P}$, denoted $d_{\mathcal{P}}(v)$, is the degree (with multiplicity, and a loop contributes two) of v in $H_{\mathcal{P}}$.

Given an F-path $P = ve_1v_1...e_tv_t$, the *ray of P from v*, denoted by $P_{v|F}$, is the longest subwalk $ve_1v_1...e_sv_s$ (possibly of length 0) of P such that $e_1, ..., e_s \in F$. There are several remarks. First, every ray is a path. Second, each F-path P has exactly two rays; these rays are identical to P if P is covered, and are edge-disjoint if P is uncovered. Third, if P is a *closed* (obviously uncovered) F-path from v to v, then both of its rays are from v. We now would like to measure the conflict between two rays. We first agree that two rays of the same F-path do *not conflict* each other, even if they may intersect at some vertex. Two rays $P_{v|F}$ and $P'_{v|F}$ (with $P \neq P'$) issued from some vertex v are *conflicting* if $P_{v|F}$ and $P'_{v|F}$ also intersect at some vertex different from v. For every $v \in V$, let $\mathcal{P}(v)$ be the set of F-paths in $\mathcal{P}$ containing v as an endpoint, and $\mathcal{P}(v|F)$ be the set of rays from v of F-paths in $\mathcal{P}$, where a closed F-path with endpoint v contributes two rays. We define the *conflict ratio* of v in $\mathcal{P}$ as

$$\mathrm{conf}_{\mathcal{P}}(v|F) := \frac{\max_{w \neq v} \left|\{P_{v|F} \in \mathcal{P}(v|F) : w \in P_{v|F}\}\right|}{d_{\mathcal{P}}(v)}.$$

We denote the *conflict ratio* of $\mathcal{P}$ by $\mathrm{conf}(\mathcal{P}|F) := \max_v \mathrm{conf}_{\mathcal{P}}(v|F)$. We always have $\mathrm{conf}(\mathcal{P}|F) \leqslant 1$ since $\left|\mathcal{P}(v|F)\right| = d_{\mathcal{P}}(v)$. Note that, when $\mathcal{P}$ is a collection, the definitions of $\mathcal{P}(v)$ and conflict ratio coincide with the definitions given in the beginning of Section 3.2.

The the proof of Theorem 4.6 is similar to the proof of Theorem 4.5 but more involved. Let us first prove an extension of Theorem 3.10 to F-decompositions. By saying a ray of $\mathcal{P}$, we mean a ray of some F-path of $\mathcal{P}$.

LEMMA 4.13

Let ℓ be a positive integer, and $\varepsilon > 0$ sufficiently small. There is an integer $L'_{\ell,\varepsilon}$ such that for every multigraph G with minimum degree at least $L'_{\ell,\varepsilon}$ and every simple

subgraph (V, F) of G, there is an F-decomposition $\mathcal{P}$ of G satisfying:

- *Every ray of $\mathcal{P}$ has length at most $\ell + 1$.*
- *Every covered F-path of $\mathcal{P}$ has length at least ℓ.*
- $\mathrm{conf}(\mathcal{P}|F) \leqslant 1/4(\ell + 9)$.
- $(1-\varepsilon)d_G(v) \leqslant \ell d_{\mathcal{P}}(v) \leqslant (1+2\varepsilon)d_G(v)$ *for every vertex v.*

Proof. Set $L'_{\ell,\varepsilon} = \max(L_{\ell,\varepsilon}, 2\ell/\varepsilon)$, where $L_{\ell,\varepsilon}$ is the constant of Theorem 3.10. We call all edges of $\overline{F} = E \backslash F$ *dummy* (note that a dummy edge may be a loop). The main idea is to replace every dummy edge by a pair of edges linking endpoints of the dummy edge to a big clique in order to obtain a simple graph to apply Theorem 3.10. For every dummy edge $e = v_{e,1}v_{e,2}$, we create a set of $L_{\ell,\varepsilon} + 1$ new vertices $X_e = \{x_{e,1}, ..., x_{e,L_{\ell,\varepsilon}+1}\}$. Let $E_e = \{x_{e,i}x_{e,j} : i \neq j\} \cup \{v_{e,1}x_{e,1}, v_{e,2}x_{e,2}\}$. Let G' be the multigraph with vertex set $\bigcup_{e\in\overline{F}} X_e \cup V$ and edge set $E' = \bigcup_{e\in\overline{F}} E_e \cup F$. It is immediate that G' is simple and $d_G(v) = d_{G'}(v)$ for every $v \in V$, and so G' has minimum degree at least L. Therefore G' admits a decomposition $\mathcal{P}'$ satisfying Theorem 3.10.

For every dummy edge e and every $i = 1, 2$, let $P'_{e,i}$ be the path of $\mathcal{P}'$ containing $v_{e,i}x_{e,i}$. We denote by $P_{i,j}$ the longest possible subwalk of $P'_{e,i}$ such that $P'_{e,i} = ...x_{e,i}v_{e,i}P_{e,i}...$ and all vertices of $P_{e,i}$ belong to V. If $P_{e,i}$ reach the end of $P'_{e,i}$, we call $P_{e,i}$ an *end-segment*; otherwise, we call it a *middle-segment*. The reader may see here the similarity between end-segments and rays. Clearly, if $P_{e,i}$ is a middle-segment, then $P'_{e,i} = ...x_{e,i}v_{e,i}P_{e,i}v_{e',j}x_{e',j}...$ for some dummy edge e' and $j \in \{1, 2\}$ since $P'_{e,i}$ leaves V right after finishing $P_{e,i}$. Note also that the lengths of end-segments and middle-segments are at most $\ell + 1$ and possibly 0.

For every dummy edge e and every $i = 1, 2$, we remove X_e and E_e, and concatenate $P_{e,i}$ with e at $v_{e,i}$. After this process, we obtain a family of walks, in which each walk lies in one of the following types:

(1) An uncovered F-path $P = P_1e_1P_2...e_{t-1}P_t$ with dummy edges $e_1, ..., e_{t-1}$, end-segments P_1 and P_t, and middle-segments $P_2, ..., P_{t-1}$. Note that the two end-segments are the rays of this uncovered F-path.

(2) A circuit without endpoint, consisting of middle-segments alternate with dummy edges but no end-segments.

Let $\mathcal{P}_1$ be the set of all the walks of Type (1) together with all paths of $\mathcal{P}'$ containing only vertices of V, and $\mathcal{P}_2$ be the set of all circuits of Type (2). Note that $\mathcal{P}_1$ is an F-collection of G, and every edge of G belongs to exactly one F-path $\mathcal{P}_1$ or one circuit of $\mathcal{P}_2$. The method of concatenating ensures that for every $v \in V$, the number of rays from v in $\mathcal{P}_1$ is equal to number of paths with endpoint v in $\mathcal{P}'$. This gives

$d_{\mathcal{P}_1}(v) = d_{\mathcal{P}'}(v)$. Besides, each ray of $\mathcal{P}_1$ is the end-segment of some path of $\mathcal{P}'$. Therefore two rays of $\mathcal{P}_1$ are conflicting only if their corresponding paths in $\mathcal{P}'$ are conflicting. Thus all of the following hold true:

- Every ray of $\mathcal{P}_1$ has length at most $\ell + 1$, since it is either a path or an end-segment of some path of $\mathcal{P}'$.
- Every covered F-path of $\mathcal{P}_1$ has length at least ℓ, since it is a path of $\mathcal{P}'$.
- $\mathrm{conf}_{\mathcal{P}_1}(v|F) \leqslant \mathrm{conf}_{\mathcal{P}'}(v) \leqslant 1/4(\ell + 10)$ for every vertex v.
- $(1-\varepsilon)d_G(v) \leqslant \ell d_{\mathcal{P}_1}(v) \leqslant (1+\varepsilon)d_G(v)$ for every v since $d_{\mathcal{P}_1}(v) = d_{\mathcal{P}'}(v)$.

We now turn our attention to $\mathcal{P}_2$. Every circuit $C \in \mathcal{P}_2$ contains at least one dummy edge. We *associate* C with some vertex v such that v is the endpoint of some dummy edge of C. For every $v \in V$, let $C_1, ..., C_t$ be the circuits (if any) associated with v, where every $C_s = ve_sW_sv$ with dummy edge e_s. Let $\widehat{P}_v = ve_1W_1ve_2W_2...ve_tW_tv$ be the walk starting and ending at v obtained by concatenating all C_s in that fashion. Clearly, $\widehat{P}_v$ is an uncovered F-path, of which one ray is v (length 0) and another ray is W_t, a middle-segment of length at most $\ell + 1$. Note that for every v, we have at most one such $\widehat{P}_v$. Let $\widehat{\mathcal{P}}_2 = \{\widehat{P}_v : v \in V\}$. Then $\widehat{\mathcal{P}}_2$ is an F-collection of G and $U_{\mathcal{P}_1} \cup U_{\widehat{\mathcal{P}}_2} = G$. Hence $\mathcal{P} = \mathcal{P}_1 \cup \widehat{\mathcal{P}}_2$ is an F-decomposition of G. Then every ray of $\mathcal{P}$ has length at most $\ell + 1$, and every covered F-path of $\mathcal{P}$ has length at least ℓ.

For every v, the number of rays from v of $\mathcal{P}$ is at most the number of rays from v of $\mathcal{P}_1$ plus two (two rays of $\widehat{P}_v$ if it exists). Hence $d_{\mathcal{P}_1}(v) \leqslant d_{\mathcal{P}}(v) \leqslant d_{\mathcal{P}_1}(v) + 2$, and so by definition of conflict ratio, we have

$$\begin{aligned}
\mathrm{conf}_{\mathcal{P}}(v|F) &\leqslant \frac{d_{\mathcal{P}_1}(v)\mathrm{conf}_{\mathcal{P}_1}(v|F) + 2}{d_{\mathcal{P}}(v)} \\
&\leqslant \mathrm{conf}_{\mathcal{P}_1}(v|F) + \frac{2}{d_{\mathcal{P}}(v)} \\
&\leqslant \frac{1}{4(\ell+10)} + \frac{2}{d_{\mathcal{P}}(v)} \\
&\leqslant \frac{1}{4(\ell+9)}.
\end{aligned}$$

Finally, we have $(1-\varepsilon)d_G(v) \leqslant \ell d_{\mathcal{P}_1}(v) \leqslant \ell d_{\mathcal{P}}(v)$. And since $L'_{\ell,\varepsilon} \geqslant 2\ell/\varepsilon$, we have $\ell d_{\mathcal{P}}(v) \leqslant \ell(d_{\mathcal{P}}(v) + 2) \leqslant (1+2\varepsilon)d_G(v)$. The proof is complete. □

Lemma 4.13 gives us a good F-decomposition $\mathcal{P}$ of G. We wish to concatenate the F-paths of $\mathcal{P}$ to an Eulerian tour. If $H_{\mathcal{P}}$ has an Eulerian tour, we naturally obtain an Eulerian tour of G by replacing each edge of $H_{\mathcal{P}}$ by its corresponding F-path of $\mathcal{P}$. Hence the goal is achieving the connectivity of $H_{\mathcal{P}}$, which immediately yields Eulerianity thank to the fact that every vertex of $H_{\mathcal{P}}$ has even degree.

LEMMA 4.14

Given a positive integer ℓ, there is an integer L'_ℓ such that for every 4-edge-connected Eulerian multigraph G with minimum degree at least $100L'_\ell$ and every simple subgraph (V, F) of G, there is an F-decomposition $\mathcal{P}$ of G satisfying:

- *Every ray of $\mathcal{P}$ has length at most $\ell + 3$.*
- *Every covered F-path of $\mathcal{P}$ has length at least ℓ.*
- $\mathrm{conf}(\mathcal{P}|F) \leqslant 1/2(\ell + 9)$.
- *$H_\mathcal{P}$ is Eulerian and spans $V(G)$.*

Proof. Let us first outline the proof. We wish to obtain connectivity of $\mathcal{P}$. To this end, we decompose G into a collection $\mathcal{P}_0$ satisfying Lemma 3.11 and two F-collections $\mathcal{P}_1$ and $\mathcal{P}_2$ satisfying Lemma 4.13. Then we use $\mathcal{P}_0$, which contains only paths of short length, to extend F-paths of $\mathcal{P}_1$ obtaining a new F-collection $\mathcal{P}'_1$ such that $H_{\mathcal{P}'_1}$ is connected. Finally, we merge $\mathcal{P}_1$ with $\mathcal{P}_2$ to obtain $\mathcal{P}$, which inherits connectivity from $\mathcal{P}_1$ and low conflict ratio from $\mathcal{P}_2$.

Because G is 4-edge-connected, by Proposition 3.6, there is an orientation D of G such that D is 2-arc-strong and $|d^+_D(v) - d^-_D(v)| \leqslant 1$ for every v. Applying Proposition 3.7 to D with an arbitrary vertex z gives us two arc-disjoint out-arborescences, T_1, T_2, rooted at z. Each vertex v has indegree at most 1 in each T_i (z has indegree 0). This gives $d_{T_1 \cup T_2}(v) \leqslant d^+_D(v) + 2 \leqslant d_G(v)/2 + 3$ for every vertex v since $|d^+_D(v) - d^-_D(v)| \leqslant 1$. Because $T_1 \cup T_2$ is loopless and 2-edge-connected, we obtain a collection $\mathcal{P}_0$ on $T_1 \cup T_2$ satisfying Lemma 3.11.

Let $G' = G \backslash U_{\mathcal{P}_0}$. Then $d_{U_{\mathcal{P}_0}}(v) \leqslant d_{T_1 \cup T_2}(v) \leqslant d_G(v)/2 + 3$, and so G' has minimum degree at least $100\ell L'_\ell/2 - 3 \geqslant 48\ell L'_\ell$. By Proposition 3.3, G' has an improper coloring by 45ℓ colors such that $|d_i(v) - d_j(v)| \leqslant 4$ for every vertex v and every pair of colors $i \neq j$. Let G_1 be the subgraph of G' with edge set of the first color, and $G_2 = G' \backslash G_1$. Thus

$$d_{G_1}(v) \leqslant \frac{1}{45\ell - 1} d_{G_2}(v) + 4 \leqslant \frac{d_{G_2}(v)}{40\ell}.$$

The minimum degrees of both G_1 and G_2 are at least $48\ell L'_\ell/45\ell - 4 \geqslant L'_\ell$. Therefore there are F-decompositions $\mathcal{P}_1$ of G_1 and $\mathcal{P}_2$ of G_2, both satisfying Lemma 4.13. Hence

$$d_{\mathcal{P}_1}(v) \leqslant \frac{1 + 2\varepsilon}{\ell} d_{G_1}(v) \leqslant \frac{1 + 2\varepsilon}{40\ell^2} d_{G_2}(v) \leqslant \frac{1 + 2\varepsilon}{40\ell(1 - \varepsilon)} d_{\mathcal{P}_2}(v),$$

for every vertex v, with an arbitrary small parameter ε. Set ε small enough (i.e. set L'_ℓ high enough) such that for every v,

$$d_{\mathcal{P}_1}(v) \leqslant \frac{1}{4(\ell + 9)} d_{\mathcal{P}_2}(v) - 3. \tag{4.1}$$

We now turn our attention to the collection $\mathcal{P}_0$ and the subcubic spanning tree $H_{\mathcal{P}_0}$. Let us consider $H_{\mathcal{P}_0}$ as a tree rooted at an arbitrary vertex z. In the following claim, we collect two private F-paths in $\mathcal{P}_1$ for each path in $\mathcal{P}_0$ for the process of concatenating later on.

CLAIM 4.15

For every path $P \in \mathcal{P}_0$ with endpoints say u, v where v is the parent of u in $H_{\mathcal{P}_0}$, there are two F-paths of $\mathcal{P}_1(v)$, named $g_1(P)$ and $g_2(P)$, such that their rays from v do not conflict with P (if $g_i(P)$ is closed, one of its rays satisfying that condition is sufficient). Furthermore, $g_i(P) \neq g_j(P')$ for any $(i, P) \neq (j, P')$.

Proof. We first apply Proposition 3.6 to have an orientation D of $H_{\mathcal{P}_1}$ such that $|d_D^-(v) - d_D^+(v)| \leqslant 1$. This orientation yields a natural orientation of F-paths of $\mathcal{P}_1$. We denote by $\mathcal{P}_1^+(v|F)$ the set of rays from v of $\mathcal{P}$ leaving v with respect to D. Note that each closed F-path at v contributes with exactly one ray to $\mathcal{P}_1^+(v|F)$. This gives $|\mathcal{P}_1^+(v|F)| \geqslant d_{\mathcal{P}_1}(v)/2 - 1$.

Since $H_{\mathcal{P}_0}$ is subcubic, there are at most 3 paths of $\mathcal{P}_0$ with endpoint v, say P_s for $1 \leqslant s \leqslant 3$. Note that each P_s has length at most 2, and so they are incident with at most 6 vertices except v in total. Recall that $\text{conf}_{\mathcal{P}_1}(v|F) \leqslant 1/4(\ell + 9)$. For each vertex w among these 6 possible vertices, we have

$$\left|\{P_{v|F} \in \mathcal{P}_1(v|F) : w \in P_{v|F}\}\right| \leqslant \frac{d_{\mathcal{P}_1}(v)}{4(\ell+9)} \leqslant \frac{2|\mathcal{P}_1^+(v|F)| + 2}{4(\ell+9)} \leqslant \frac{|\mathcal{P}_1^+(v|F)|}{12}.$$

Hence in total there are at most $|\mathcal{P}_1^+(v|F)|/2$ rays of $\mathcal{P}_1^+(v|F)$ conflicting with some P_s. This guarantees that there are at least half of rays in $\mathcal{P}_1^+(v)$ non-conflicting with all P_s. We just pick 6 rays among them, and name the F-paths of these rays $g_i(P_s)$ arbitrarily (these F-paths are clearly pairwise distinct). Note also that $\mathcal{P}_1^+(v) \cap \mathcal{P}_1^+(v') = \emptyset$ for any $v \neq v'$, so $g_i(P) \neq g_j(P')$ for any $(i, P) \neq (j, P')$. □

We can now obtain the connectivity of H_1 by concatenating each P of $\mathcal{P}_0$ to either $g_1(P)$ or $g_2(P)$. We again call a *structured-tree* $\mathcal{T}$ a rooted tree on vertex set $\{Y_1, Y_2, ..., Y_t\}$, where $\{Y_1, Y_2, ..., Y_t\}$ is some partition of V with the following properties:

(A) For every edge Y_iY_j of $\mathcal{T}$, there is a corresponding path $v_i...v_j \in \mathcal{P}_0$, where $v_i \in Y_i$ and $v_j \in Y_j$.

(B) For every Y_i, there is an F-collection $\mathcal{R}_i$ such that $H_{\mathcal{R}_i}$ is connected and spans Y_i, and each F-path in $\mathcal{R}_i$ is either $g_1(P)$ or the concatenation of P and $g_1(P)$ for some $P \in \mathcal{P}_0$ (if Y_i contains a single vertex then $\mathcal{R}_i$ is empty).

Such structured-tree $\mathcal{T}$ clearly exists by choosing $\mathcal{T}$ equal to $H_{\mathcal{P}_0}$ rooted at z, in which each Y_i contains a single vertex, and each $\mathcal{R}_i$ is empty. Our goal is to repeatedly merge vertices of $\mathcal{T}$ until $\mathcal{T}$ is the singleton graph, which completes the process of concatenating. We consider a leaf Y_i of $\mathcal{T}$ with parent Y_j, corresponding to path $P = v_i...v_j$ of $\mathcal{P}_0$ with $v_i \in Y_i$ and $v_j \in Y_j$. Suppose that $g_1(P) = v_j...y$ and $g_2(P) = v_j...z$.

- If $y \in Y_k$ for some $k \neq i$, we concatenate P and $g_1(P)$ at v_j and get a F-path P^*. Then we merge Y_i into Y_k to form new set Y_{ik} (inheriting the position of Y_k in tree $\mathcal{T}$). Let $\mathcal{R}_{ik} = \mathcal{R}_i \cup \mathcal{R}_k \cup \{P^*\}$. Since P^* connects two vertices of $\mathcal{R}_i$ and $\mathcal{R}_k$, we have that $H_{\mathcal{R}_{ik}}$ is connected and spans Y_{ik}.
- If $y \in Y_i$, we merge Y_i to Y_j to form new set Y_{ij} (inheriting the position of Y_j in tree $\mathcal{T}$). Set $\mathcal{R}_{ij} = \mathcal{R}_i \cup \mathcal{R}_j \cup \{g_1(P)\}$. Since $g_1(P)$ connects two vertices of $\mathcal{R}_i$ and $\mathcal{R}_j$, we have that $H_{\mathcal{R}_{ij}}$ is connected and spans Y_{ij}. We also concatenate P with $g_2(P)$ at v_j to get another F-path and put it back into $\mathcal{P}_1$.

The number of vertices of $\mathcal{T}$ is reduced by 1 after each step, while $\mathcal{T}$ still satisfies both properties. Once the process is complete, we end up with a singleton $\mathcal{T}$ and an F-collection $\mathcal{R}$ such that $H_{\mathcal{R}}$ is connected and spans V. Note that $\mathcal{P}_0$ is empty at the end of the process, since exactly one path of $\mathcal{P}_0$ is used at each step. We merge $\mathcal{R}$ with $\mathcal{P}_1$ to obtain a new collection $\mathcal{P}_1'$. Consequently, $H_{\mathcal{P}_1'}$ is connected.

Let $\mathcal{P} = \mathcal{P}_1' \cup \mathcal{P}_2$. Note that $U_{\mathcal{P}} = U_{\mathcal{P}_1'} \cup U_{\mathcal{P}_2} = G$, so $\mathcal{P}$ is an F-decomposition of G and $H_{\mathcal{P}}$ is connected. The degrees of all vertices of G are even, then so are the degrees of vertices of $H_{\mathcal{P}}$, and hence $H_{\mathcal{P}}$ is Eulerian. The process of concatenating also ensures that every ray of $\mathcal{P}$ has length at most $\ell + 3$ and that every covered F-path of $\mathcal{P}$ has length at least ℓ.

It remains to prove that $\mathrm{conf}(\mathcal{P}|F) \leqslant 1/2(\ell + 9)$. In the following, by saying $\mathcal{P}_0$ or $\mathcal{P}_1$, we mean the collection before the process of concatenating. Recall that $H_{\mathcal{P}_0}$ is subcubic, so for every vertex v, the number of F-paths with endpoint v in $\mathcal{P}_1'$ is at most the number F-paths with endpoint v in $\mathcal{P}_1$ plus 3. Combining with (4.1) yields $d_{\mathcal{P}_1'}(v) \leqslant d_{\mathcal{P}_1}(v) + 3 \leqslant d_{\mathcal{P}_2}(v)/4(\ell + 9)$. Recall that $\mathrm{conf}_{\mathcal{P}_1'}(v|F) \leqslant 1$ and $\mathrm{conf}_{\mathcal{P}_2}(v|F) \leqslant 1/4(\ell + 9)$. Hence for every vertex v, by definition of conflict ratio we have

$$\begin{aligned}
\mathrm{conf}_{\mathcal{P}}(v|F) &\leqslant \frac{d_{\mathcal{P}_2}(v)\mathrm{conf}_{\mathcal{P}_2}(v|F) + d_{\mathcal{P}_1'}(v)\mathrm{conf}_{\mathcal{P}_1'}(v|F)}{d_{\mathcal{P}_2}(v) + d_{\mathcal{P}_1}(v)} \\
&< \mathrm{conf}_{\mathcal{P}_2}(v|F) + \frac{d_{\mathcal{P}_1'}(v)\mathrm{conf}_{\mathcal{P}_1'}(v|F)}{d_{\mathcal{P}_2}(v)} \\
&\leqslant \frac{1}{4(\ell+9)} + \frac{1}{4(\ell+9)} \\
&\leqslant \frac{1}{2(\ell+9)}.
\end{aligned}$$

This implies $\mathrm{conf}(\mathcal{P}|F) \leqslant 1/2(\ell + 9)$, and the lemma follows. □

The final step is concatenating F-paths of $\mathcal{P}$ to obtain a well-behaved Eulerian tour of G, which can be done thank to Proposition 3.16.

Proof of Theorem 4.6. Let $d_\ell = 100\ell L'_{\ell,\varepsilon}$ and $G' = (V, F)$. We first obtain an F-decomposition $\mathcal{P}$ of G satisfying Lemma 4.14. For every ray $P_{v|F}$ of $\mathcal{P}$, each vertex $w \in P_{v|F}$ is a conflict point between $P_{v|F}$ and at most $d_{\mathcal{P}}(v)/2(\ell+9)$ other rays. Hence the number of rays conflicting with $P_{v|F}$ is at most $(\ell+3)d_{\mathcal{P}}(v)/2(\ell+9) \leqslant d_{\mathcal{P}}(v)/2 - 2$ since $P_{v|F}$ has length at most $\ell+3$.

We wish to apply Proposition 3.16 to $H_{\mathcal{P}}$. Therefore the task now is to eliminate all loops of $H_{\mathcal{P}}$. Let $H^*_{\mathcal{P}}$ be the loopless multigraph obtained from $H_{\mathcal{P}}$ by subdividing every loop $e = vv$ into vx_e and x_ev by a new vertex x_e. We associate each vx_e and x_ev with a ray of P, where $P \in \mathcal{P}(v)$ is the corresponding F-path of e.

For every pair of incident vertex-edge (v, e) of $H^*_{\mathcal{P}}$, let $S_v(e)$ be the set of all edges of $H^*_{\mathcal{P}}$ corresponding to rays conflicting with $P_{v|F}$, where $P_{v|F}$ is ray of $\mathcal{P}$ corresponding to e. Since two rays of the same F-path are non-conflicting, we have $|S_{x_e}(e)| = 0$ for every loop e of $H^*_{\mathcal{P}}$. Hence $|S_v(e)| \leqslant d_{H^*_{\mathcal{P}}}(v)/2 - 2$ if $d_{H^*_{\mathcal{P}}}(v) \geqslant 4$ and $|S_v(e)| = 0$ if $d_{H^*_{\mathcal{P}}}(v) = 2$ for every pair of incident vertex-edge (v, e) of $H^*_{\mathcal{P}}$.

Let $\mathcal{S} = \{S_v\}_{v\in V}$, then $\mathcal{S}$ is a generalized transition system of $H^*_{\mathcal{P}}$. Proposition 3.16 asserts that $H^*_{\mathcal{P}}$ admits an Eulerian tour $\mathcal{E}_{H^*_{\mathcal{P}}}$ compatible with $\mathcal{S}$, i.e., the corresponding rays of any two consecutive edges of $\mathcal{E}_{H^*_{\mathcal{P}}}$ are non-conflicting. Clearly, vx_e and x_ev are two consecutive edges of $\mathcal{E}_{H^*_{\mathcal{P}}}$ since x_e has degree 2. We therefore naturally obtain from $\mathcal{E}_{H^*_{\mathcal{P}}}$ an Eulerian tour $\mathcal{E}_{H_{\mathcal{P}}}$ of $H_{\mathcal{P}}$ by replacing e to the segment vx_ev for every loop $e = vv$ of $H_{\mathcal{P}}$. Hence we naturally obtain from $\mathcal{E}_{H_{\mathcal{P}}}$ an Eulerian tour $\mathcal{E}$ of G by replacing every edge of $\mathcal{E}_{H_{\mathcal{P}}}$ by its corresponding F-path of $\mathcal{P}$. Note that every two consecutive (with respect to $\mathcal{E}$) rays of $\mathcal{P}$ are non-conflicting.

Let W be a segment of $\mathcal{E}$ of length at most ℓ and consisting of only edges of F. It remains to prove that W is a path. Let $P_1, P_2..., P_r$ be consecutive (with respect to $\mathcal{E}$) F-paths of $\mathcal{P}$ such that W is a subwalk of $P_1P_2...P_r$ and $W \cap P_1, W \cap P_r \neq \emptyset$. If $r \geqslant 3$ then W must contain entirely P_2. All edges of W belong to F, then so does P_2. Hence P_2 is a covered F-path of length at most $\ell - 2$, contrary to the fact that every covered F-path of $\mathcal{P}$ has length at least ℓ. If $r = 2$, note that the rays from v of P_1 and P_2 are non-conflicting, and W is a subwalk of the concatenation of these two rays. Hence W is a path. If $r = 1$ then clearly W is a path, the desired conclusion. □

Chapter 5

Additive Bases and Flows

5.1 Flows and weighted flows

In this final chapter of Part I, we will return to flows and discuss several results in flows, weighted flows and weighted orientations in graphs. All of them are derived from a special case of a conjecture by Jaeger, Linial, Payan and Tarsi [JLPT92]. Unlike previous chapters, graphs considered in this chapter *may have multiple edges but no loops* (i.e. multigraphs in previous chapters). As we mentioned before, the result 8-edge-connected graphs having a 3-flow by Thomassen [Tho12] was improved by [LTWZ13] to 6-edge-connected graphs. In fact, in [LTWZ13], they proved a more general result – Theorem 5.1, that edges of a highly edge-connected graph can be oriented so that any prescribed value modulo k at every vertex can be realized.

Recall that an *orientation* $\vec{G} = (V, \vec{E})$ of a graph $G = (V, E)$ is obtained by giving each edge of E a direction. For each edge $e \in E$, we denote in this chapter the corresponding arc of $\vec{E}$ by $\vec{e}$, and vice versa. For an integer $p \geqslant 2$, a mapping $\beta : V \to \mathbb{Z}_p$ is a *p-boundary* of G if $\sum_{v \in V} \beta(v) \equiv 0 \pmod{p}$. Given a p-boundary β of G, an orientation $\vec{G}$ of G is a *β-orientation* if $d^+_{\vec{G}}(v) - d^-_{\vec{G}}(v) \equiv \beta(v) \pmod{p}$ for every $v \in V$.

THEOREM 5.1 [LTWZ13]

For any $k \geqslant 1$, any $6k$-edge-connected graph G and any $(2k+1)$-boundary β of G, the graph G has a β-orientation.

The content of this chapter is covered in paper *Additive bases and flows in graphs*, Louis Esperet, Rémi de Joannis de Verclos, Tien-Nam Le, and Stéphan Thomassé, *SIAM Journal of Discrete Mathematics*, 32:534–542, 2018.

By setting $k = 1$ and $\beta(v) = 0$ for every v, we obtain that every 6-edge-connected graph has a 3-flow. A natural question is whether a weighted version of Theorem 5.1 exists. Given a graph $G = (V, E)$, a p-boundary β of G and a mapping $f : E \to \mathbb{Z}_p \setminus \{0\}$, an orientation $\vec{G}$ of G is called an *f-weighted β-orientation* if $\partial f(v) \equiv \beta(v) \pmod p$ for every v, where $\partial f(v) = \sum_{\vec{e} \in E^+_{\vec{G}}(v)} f(e) - \sum_{\vec{e} \in E^-_{\vec{G}}(v)} f(e)$. Note that if $f(e) \equiv 1 \pmod p$ for every edge e, an f-weighted β-orientation is precisely a β-orientation.

QUESTION 5.2

For which mapping $f : E \to \mathbb{Z}_{2k+1} \setminus \{0\}$ and $(2k+1)$-boundary β, every highly edge-connected graph G has an f-weighted β-orientation?

An immediate observation is that if we wish to have a general result of Question 5.2, it is necessary to assume that $2k + 1$ is a prime number. For instance, take G to consist of two vertices u, v with an arbitrary number of edges between u and v, consider a non-trivial divisor p of $2k + 1$, and ask for a $\mathbf{p}$-weighted $(2k + 1)$-orientation $\vec{G}$ of G (here, $\mathbf{p}$ denotes the function that maps each edge to $p \pmod{2k+1}$). Note that for any orientation, $\partial\mathbf{p}(v)$ is in the subgroup of $\mathbb{Z}_{2k+1}$ generated by p, and this subgroup does not contain $1, -1 \pmod{2k+1}$. In particular, there is no $\mathbf{p}$-weighted $(2k + 1)$-orientation of G with boundary β satisfying $\beta(u) \equiv -\beta(v) \equiv 1 \pmod{2k+1}$.

We will show later on that when k is a prime number, the answer for Question 5.2 is affirmative for every mapping f and boundary β.

THEOREM 5.3

Let $p \geqslant 3$ be a prime number and let $G = (V, E)$ be a $(6p - 8)(p - 1)$-edge-connected graph. For any mapping $f : E \to \mathbb{Z}_p \setminus \{0\}$ and any p-boundary β, G has an f-weighted β-orientation.

Theorem 5.3 is indeed a direct corollary of a special case of Jaeger-Linial-Payan-Tarsi's conjecture [JLPT92] on additive basis of the vector space $\mathbb{Z}_p^n$, which we will present and prove in the next section. We will also discuss various results on orientations and flows in $\mathbb{Z}_p$ obtained from that special case of Jaeger-Linial-Payan-Tarsi's conjecture.

5.2 Jaeger-Linial-Payan-Tarsi's conjecture

An *additive basis* B of a vector space F is a multiset of elements from F such that for all $\beta \in F$, there is a subset of B which sums to β. Let $\mathbb{Z}_p^n$ be the n-dimensional linear space over the prime field $\mathbb{Z}_p$. The following result is a simple consequence of the Cauchy-Davenport Theorem [Dav35] (see also [ANR96]).

THEOREM 5.4 [Dav35]

For any prime p, any multiset of $p-1$ non-zero elements of $\mathbb{Z}_p$ forms an additive basis of $\mathbb{Z}_p$.

This result can be rephrased as: *for $n = 1$, any family of $p-1$ linear bases of $\mathbb{Z}_p^n$ forms an additive basis of $\mathbb{Z}_p^n$.* A natural question is whether this can be extended to all integers n. Given a collection of sets $X_1, ..., X_k$, we denote by $\uplus_{i=1}^{k} X_i$ the union with repetitions of $X_1, ..., X_k$. Jaeger, Linial, Payan and Tarsi [JLPT92] conjectured the following, which as we will see is a generalization of important results regarding flows in graphs.

CONJECTURE 5.5 [JLPT92]

For every prime number p, there is a constant $c(p)$ such that for any $t \geqslant c(p)$ linear bases $B_1, ..., B_t$ of $\mathbb{Z}_p^n$, the union $\uplus_{s=1}^{t} B_s$ forms an additive basis of $\mathbb{Z}_p^n$.

Alon, Linial and Meshulam [ALM91] proved a weaker version of Conjecture 5.5, that the union of any $p\lceil \log n \rceil$ linear bases of $\mathbb{Z}_p^n$ contains an additive basis of $\mathbb{Z}_p^n$ (note that unlike Conjecture 5.5, their bound depends on n).

Let us call the *support* of a vector $x = (x_1, \ldots, x_n) \in \mathbb{Z}_p^n$ the set of indices i such that $x_i \neq 0$, and the *shadow* of a vector x the (unordered) multiset of non-zero entries of x. Note that sizes of the support and of the shadow of a vector are equal. We will prove, as the main result in this chapter, that Conjecture 5.5 holds if the support of each vector has size at most two.

THEOREM 5.6

Let $p \geqslant 3$ be a prime number. For some integer $\ell \geqslant 1$, consider $t \geqslant 8\ell(3p-4) + p - 2$ linear bases $B_1, ..., B_t$ of $\mathbb{Z}_p^n$, such that the support of each vector has size at most 2, and at most ℓ different shadows of size 2 appear among the vectors of $\mathcal{B} = \uplus_{s=1}^{t} B_s$. Then $\mathcal{B}$ forms an additive basis of $\mathbb{Z}_p^n$.

While Theorem 5.3, a generalization of Theorem 5.1, can be implied from Theorem Theorem 5.6, Theorem 5.6 will indeed be proved in using Theorem 5.1.

5.3 Corollaries on additive bases and flows

Corollaries on additive bases

Observe that the number of possibilities for an (unordered) multiset of $\mathbb{Z}_p \setminus \{0\}$ of size 2 is $\binom{p-1}{2} + p - 1 = \binom{p}{2}$. As a consequence, Theorem 5.6 has the following immediate

corollary.

COROLLARY 5.7

Let $p \geqslant 3$ be a prime number. For any $t \geqslant 8\binom{p}{2}(3p-4)+p-2$ linear bases $B_1, ..., B_t$ of $\mathbb{Z}_p^n$ such that the support of each vector has size at most 2, $\biguplus_{s=1}^{t} B_s$ forms an additive basis of $\mathbb{Z}_p^n$.

Another interesting consequence of Theorem 5.6 concerns the linear subspace $(\mathbb{Z}_p^n)_0$ of vectors of $\mathbb{Z}_p^n$ whose entries sum to $0 \pmod p$.

COROLLARY 5.8

Let $p \geqslant 3$ be a prime number. For any $t \geqslant 4(p-1)(3p-4)+p-2$ linear bases $B_1, ..., B_t$ of $(\mathbb{Z}_p^n)_0$ such that the support of each vector has size at most 2, $\biguplus_{s=1}^{t} B_s$ forms an additive basis of $(\mathbb{Z}_p^n)_0$.

Proof. Note that for any $1 \leqslant s \leqslant t$, the linear basis B_s consists of $n-1$ vectors, each of which has a support of size 2, and the two elements of the shadow sum to $0 \pmod p$. In particular, at most $\frac{p-1}{2}$ different shadows appear among the vectors of the linear bases $B_1, ..., B_t$. It is convenient to view each B_s as a matrix in which the elements of the basis are column vectors. For each $1 \leqslant s \leqslant t$, let B'_s be obtained from B_s by deleting the last row. It is easy to see that B'_s is a linear basis of $\mathbb{Z}_p^{n-1}$. Moreover, at most $\frac{p-1}{2}$ different shadows of size 2 appear among the vectors of the linear bases $B'_1, ..., B'_t$ (note that the removal of the last row may have created vectors with shadows of size 1). In particular, it follows from Theorem 5.6 that for any vector $\beta = (\beta_1, \ldots, \beta_n) \in (\mathbb{Z}_p^n)_0$, the vector $(\beta_1, \ldots, \beta_{n-1}) \in \mathbb{Z}_p^{n-1}$ can be written as a sum of a subset of elements of $\biguplus_{s=1}^{t} B'_s$. Clearly, the corresponding subset of elements of $\biguplus_{s=1}^{t} B_s$ sums to β. This concludes the proof of Corollary 5.8. □

Corollaries on $\mathbb{Z}_p$-flows

Theorem 5.3 turns out to be equivalent to the following seemingly more general result. Assume that we are given a directed graph $\vec{G} = (V, \vec{E})$ and a p-boundary β. A *p-flow with boundary β* in $\vec{G}$ is a mapping $f : \vec{E} \to \mathbb{Z}_p$ such that $\partial f(v) \equiv \beta(v) \pmod p$ for every v. In other words, f is a p-flow with boundary β in $\vec{G} = (V, \vec{E})$ if and only if $\vec{G}$ is an f-weighted β-orientation of its underlying non-oriented graph $G = (V, E)$, where f is extended from $\vec{E}$ to E in the natural way (i.e. for each $e \in E$, $f(e) := f(\vec{e})$).

In the rest of this chapter we will say that a directed graph $\vec{G}$ is *t-edge-connected* if its underlying graph, denoted by G, is t-edge-connected.

THEOREM 5.9

Let $p \geqslant 3$ be a prime number and let $\vec{G} = (V, \vec{E})$ be a directed $(6p-8)(p-1)$-edge-connected graph. For any arc $\vec{e} \in \vec{E}$, let $L(\vec{e})$ be a pair of distinct elements of $\mathbb{Z}_p$. Then for every p-boundary β, $\vec{G}$ has a $\mathbb{Z}_p$-flow f with boundary β such that for any $\vec{e} \in \vec{E}$, $f(\vec{e}) \in L(\vec{e})$.

To see that Theorem 5.9 implies Theorem 5.3, simply fix an arbitrary orientation of G and set $L(\vec{e}) = \{f(e), -f(e)\}$ for each arc $\vec{e}$. We now prove that Theorem 5.3 implies Theorem 5.9. We actually prove a slightly stronger statement (holding in $\mathbb{Z}_{2k+1}$ for any integer $k \geqslant 1$).

LEMMA 5.10

Let $k \geqslant 1$ be an integer, and let $\vec{G} = (V, \vec{E})$ be a directed graph such that the underlying non-oriented graph G has an f-weighted β-orientation for any mapping $f : E \to \mathbb{Z}_{2k+1} \setminus \{0\}$ and any $(2k+1)$-boundary β. For every arc $\vec{e} \in \vec{E}$, let $L(\vec{e})$ be a pair of distinct elements of $\mathbb{Z}_{2k+1}$. Then for every $(2k+1)$-boundary β, $\vec{G}$ has a $(2k+1)$-flow g with boundary β such that $g(\vec{e}) \in L(\vec{e})$ for every $\vec{e}$.

Proof. Let β be a $(2k+1)$-boundary of $\vec{G}$. Consider a single arc $\vec{e} = (u, v)$ of $\vec{G}$. Choosing one of the two values of $L(\vec{e})$, say a or b, will either add a to $\partial g(u)$ and subtract a from $\partial g(v)$, or add b to $\partial g(u)$ and subtract b from $\partial g(v)$. Note that 2 and $2k+1$ are relatively prime, so the element 2^{-1} is well-defined in $\mathbb{Z}_{2k+1}$. If we now add $2^{-1}(a+b)$ to $\beta(v)$ and subtract $2^{-1}(a+b)$ from $\beta(u)$, the earlier choice is equivalent to choosing between the two following options: adding $2^{-1}(a-b)$ to $\partial g(u)$ and subtracting $2^{-1}(a-b)$ from $\partial g(v)$, or adding $2^{-1}(b-a)$ to $\partial g(u)$ and subtracting $2^{-1}(b-a)$ from $\partial g(v)$. This is equivalent to choosing an orientation for an edge of weight $2^{-1}(a-b)$. It follows that finding a $(2k+1)$-flow g with boundary β such that for any $\vec{e} \in \vec{E}$, $g(\vec{e}) \in L(\vec{e})$ is equivalent to finding an f-weighted β'-orientation for some other $(2k+1)$-boundary β' of G, where the weight $f(e)$ of each edge e is 2^{-1} times the difference between the two elements of $L(\vec{e})$. □

We now consider the case where $L(\vec{e}) = \{0, 1\}$ for every arc $\vec{e} \in \vec{E}$. Let $f_{2^{-1}} : \vec{E} \to \mathbb{Z}_{2k+1}$ denote the function that maps each arc $\vec{e}$ to $2^{-1} \pmod{2k+1}$. The same argument as in the proof of Lemma 5.10 implies that if G has an $f_{2^{-1}}$-weighted β-orientation for every $(2k+1)$-boundary β, then for every $(2k+1)$-boundary β, the digraph $\vec{G}$ has a $(2k+1)$-flow f with boundary β such that $f(\vec{e}) \in L(\vec{e})$ for every $\vec{e}$.

The following is a simple corollary of Theorem 5.1.

COROLLARY 5.11

> *Let $\ell \geqslant 1$ be an odd integer and let $k \geqslant 1$ be relatively prime with ℓ. Let $G = (V, E)$ be a $(3\ell - 3)$-edge-connected graph, and let $\mathbf{k} : E \to \mathbb{Z}_\ell$ be the mapping that assigns $k \pmod{\ell}$ to each edge $e \in E$. Then for any ℓ-boundary β, G has a $\mathbf{k}$-weighted β-orientation.*

Proof. Observe that $\beta' = k^{-1} \cdot \beta$ is a ℓ-boundary (k^{-1} is well defined in $\mathbb{Z}_\ell$). It follows from Theorem 5.1 that G has a β'-orientation. Note that this corresponds to a $\mathbf{k}$-weighted β-orientation of G, as desired. □

As a consequence, the following is an equivalent version of Theorem 5.1 (see also [JLPT92, LL17]).

THEOREM 5.12

> *Let $k \geqslant 1$ be an integer and let $\vec{G} = (V, \vec{E})$ be a directed $6k$-edge-connected graph. Then for every $(2k+1)$-boundary β, $\vec{G}$ has a $(2k+1)$-flow f with boundary β such that $f(\vec{E}) \in \{0, 1\} \pmod{2k+1}$.*

5.4 Additive bases in matrices with small support

This section is devoted to prove Theorem 5.6 and show how it implies Theorem 5.3. We first recall the following (weak form of a) classical result by Mader (see [Die05], Theorem 1.4.3). Note that the *average degree* of a graph G is $\frac{|E(G)|}{|V(G)|}$.

LEMMA 5.13

> *Given an integer $k \geqslant 1$, if $G = (V, E)$ is a graph with average degree at least $4k$, then there is a subset X of V such that $|X| > 1$ and $G[X]$ is $(k+1)$-edge-connected.*

We will also need the following result of Thomassen [Tho14], which is a simple consequence of Theorem 5.1.

THEOREM 5.14 [Tho14]

> *Let $k \geqslant 3$ be an odd integer, $G = (V_1, V_2, E)$ be a bipartite graph, and $f : V_1 \cup V_2 \to \mathbb{Z}_k$ be a mapping satisfying $\sum_{v \in V_1} f(v) \equiv \sum_{v \in V_2} f(v) \pmod{k}$. If G is $(3k-3)$-edge-connected, then G has a spanning subgraph H such that for any $v \in V$, $d_H(v) \equiv f(v) \pmod{k}$.*

Let G be a graph, and let X and Y be two disjoint subsets of vertices of G. The set of edges of G with one endpoint in X and the other in Y is denoted by $E(X, Y)$.

We are now ready to prove Theorem 5.6.

Proof of Theorem 5.6. We proceed by induction on n. For $n = 1$, this is a direct consequence of Theorem 5.4, so suppose that $n \geqslant 2$. Each basis B_s can be considered as an $n \times n$ matrix where each column is a vector with support of size at most 2. Let $\mathcal{B} = \biguplus_{i=1}^{t} B_i$.

For $1 \leqslant i \leqslant n$, a vector is called an *i-vector* if its support is the singleton $\{i\}$ (in other words, the i-th entry is non-zero and all the other entries are zero). Suppose that for some $1 \leqslant i \leqslant n$, $\mathcal{B}$ contains at least $p-1$ i-vectors. Let $\mathcal{C}$ be the set of i-vectors of $\mathcal{B}$. Clearly, each basis contains at most one i-vector. For every B_s, let B'_s be the matrix obtained from B_s by removing its i-vector (if any) and the i^{th} row. Clearly B'_s is or contains a basis of $\mathbb{Z}_p^{n-1}$. By induction hypothesis, $\biguplus_{s=1}^{t} B'_s$ forms an additive basis of $\mathbb{Z}_p^{n-1}$. In other words, for any vector $\beta = (\beta_1, ..., \beta_i, ..., \beta_n) \in \mathbb{Z}_p^n$, there is a subset Y_1 of $\mathcal{B} \setminus \mathcal{C}$ which sums to $(\beta_1, ..., \hat{\beta}_i, .., \beta_n)$ for some $\hat{\beta}_i$. Since $|\mathcal{C}| \geqslant p-1$, it follows from Theorem 5.4 that there is a subset Y_2 of $\mathcal{C}$ which sums to $(0, ..., \beta_i - \hat{\beta}_i, .., 0)$. Hence $Y_1 \cup Y_2$ sums to β.

Thus we can suppose that there are at most $p-2$ i-vectors for every i. Then there are at least $8\ell(3p-4)n$ vectors with a support of size 2 in $\mathcal{B}$. Since there are at most ℓ distinct shadows of size 2 in $\mathcal{B}$, there are at least $8(3p-4)n$ vectors with the same (unordered) shadow of size 2, say $\{a_1, a_2\}$ (recall that shadows are multisets, so a_1 and a_2 might coincide).

Let G be the graph (recall that graphs in this chapter are allowed to have multiple edges) with vertex set $V = \{v_1, ..., v_n\}$ and edge set E, where edges $v_i v_j$ are in one-to-one correspondence with vectors of $\mathcal{B}$ with support $\{i, j\}$ and shadow $\{a_1, a_2\}$. Then G contains at least $8(3p-4)n$ edges.

We now consider a random partition of V into 2 sets V_1, V_2 (by assigning each vertex of V uniformly at random to one of the sets V_k, $k = 1, 2$). Let $e = v_i v_j$ be some edge of G. Recall that e corresponds to some vector with only two non-zero entries, say without loss of generality a_1 at i^{th} index and a_2 at j^{th} index. The probability that v_i is assigned to V_1 and v_j is assigned to V_2 is at least $\frac{1}{4}$. As a consequence, there is a partition of V into 2 sets V_1, V_2 and a subset $E' \subseteq E(V_1, V_2)$ of at least $8(3p-4)n/4 = 2(3p-4)n$ edges such that for every $e \in E'$, the vector of $\mathcal{B}$ corresponding with e has entry a_1 (resp. a_2) at the index associated to the endpoint of e in V_1 (resp. V_2).

Since the graph $G' = (V, E')$ has average degree at least $4(3p-4)$, it follows from Lemma 5.13 that there is a set $X \subseteq V$ of at least 2 vertices, such that $G'[X]$ is $(3p-3)$-edge-connected. Set $H = G'[X]$ and F the edge set of H. Note that H is bipartite with bipartition $X_1 = X \cap V_1$ and $X_2 = X \cap V_2$.

For each integer $1 \leqslant s \leqslant t$, let B_s^* be the matrix obtained from B_s by doing the following: for each vertex v_i in X_1 (resp. X_2), we multiply all the elements of the i^{th} row of B_s by a_1^{-1} (resp. $-a_2^{-1}$), noting that all the operations are performed in $\mathbb{Z}_p$. Let $\mathcal{B}^* = \biguplus_{s=1}^{t} B_s^*$. Note that each vector of $\mathcal{B}^*$ corresponding to some edge $e \in F$ has shadow $\{1, -1\}$ (1 is the entry indexed by the endpoint of e in X_1 and -1 is the entry indexed by the endpoint of e in X_2). It is easy to verify the following.

- Each B_s^* is a linear basis of $\mathbb{Z}_p^n$.
- $\mathcal{B}$ is an additive basis if and only if $\mathcal{B}^*$ is an additive basis.

Hence it suffices to prove that $\mathcal{B}^*$ is an additive basis.

Without loss of generality, suppose that $X = \{v_m, ..., v_n\}$ for some $m \leqslant n-1$. By *contracting* k rows of a matrix, we mean deleting these k rows and adding a new row consisting of the sum of the k rows. For each $1 \leqslant s \leqslant t$, let B'_s be the matrix of m rows obtained from B_s^* by contracting all $m^{th}, (m+1)^{th}, ..., n^{th}$ rows. Note that the operation of contracting k rows decreases the rank of the matrix by at most $k-1$ (since it is the same as replacing one of the rows by the sum of the k rows, which preserves the rank, and then deleting the $k-1$ other rows). Let $\mathcal{B}' = \biguplus_{s=1}^{t} B'_s$. Since each B_s^* is a linear basis of $\mathbb{Z}_p^n$, each B'_s has rank at least m and therefore contains a basis of $\mathbb{Z}_p^m$. Hence, by induction hypothesis, $\mathcal{B}' \setminus \mathcal{B}'_0$ is an additive basis of $\mathbb{Z}_p^m$, where $\mathcal{B}'_0$ is the set of all columns with empty support in $\mathcal{B}'$. For every $\beta = (\beta_1, ..., \beta_n) \in \mathbb{Z}_p^n$, let $\beta' = (\beta_1, ..., \beta_{m-1}, \sum_{i=m}^n \beta_i) \in \mathbb{Z}_p^m$. Then there is a subset Y' of $\mathcal{B}' \setminus \mathcal{B}'_0$ which sums to β'. Let Y^* and $\mathcal{B}_0^*$ be the subsets of $\mathcal{B}^*$ corresponding to Y' and $\mathcal{B}'_0$, respectively. Then Y^* sums to some $\hat{\beta} = (\beta_1, ..., \beta_{m-1}, \hat{\beta}_m, ..., \hat{\beta}_n)$, where $\sum_{i=m}^n \hat{\beta}_i \equiv \sum_{i=m}^n \beta_i \pmod p$.

Recall that for each edge $e \in F$, the corresponding vector in $\mathcal{B}^*$ has precisely two non-zero entries, $(1, -1)$, each with index in X. Hence the vector corresponding to each $e \in F$ in $\mathcal{B}'$ has empty support. Thus the set of vectors in $\mathcal{B}^*$ corresponding to the edge set F is a subset of $\mathcal{B}_0^*$, which is disjoint from Y.

For each $v_i \in X_1$, let $\beta_X(v_i) = \beta_i - \hat{\beta}_i$, and for each $v_i \in X_2$, let $\beta_X(v_i) = \hat{\beta}_i - \beta_i$. Since $\sum_{i=m}^n \hat{\beta}_i \equiv \sum_{i=m}^n \beta_i \pmod p$, we have $\sum_{v_i \in X \cap V_1} \beta_X(v_i) = \sum_{v_i \in X \cap V_2} \beta_X(v_i)$. Since H is $(3p-3)$-edge-connected, it follows from Theorem 5.14 that there is a subset $F' \subseteq F$ such that, in the graph (X, F'), each vertex $v_i \in X_1$ has degree $\beta_i - \hat{\beta}_i \pmod p$ and each vertex $v_i \in X_2$ has degree $\hat{\beta}_i - \beta_i \pmod p$. Therefore, F' corresponds to a subset Z^* of vectors of $\mathcal{B}_0^*$, summing to $(0, \ldots, 0, \beta_m - \hat{\beta}_m, \ldots, \beta_n - \hat{\beta}_n)$. Then $Y^* \cup Z^*$ sums to β. It follows that $\mathcal{B}^*$ is an additive basis of $\mathbb{Z}_p^n$, and so is $\mathcal{B}$. This completes the proof. □

Proofs of Theorem 5.3

We will complete this chapter with two proofs of Theorem 5.3. The first one is a direct application of Corollary 5.8, but requires a stronger assumption on the edge-connectivity of G ($24p^2 - 54p + 28$ instead of $6p^2 - 14p + 8$ for the second proof).

First proof of Theorem 5.3. We fix some arbitrary orientation $\vec{G} = (V, \vec{E})$ of G and denote the vertices of G by $v_1, \ldots, v_n$. The number of edges of G is denoted by m. For each arc $\vec{e} = (v_i, v_j)$ of $\vec{G}$, we associate $\vec{e}$ to a vector $x_e \in (\mathbb{Z}_p^n)_0$ in which the i^{th}-entry is equal to $f(e) \pmod p$, the j^{th}-entry is equal to $-f(e) \pmod p$ and all the remaining entries are equal to $0 \pmod p$.

Let us consider the following statements.

(a) For each p-boundary β, there is an f-weighted β-orientation of G.

(b) For each p-boundary β there is a vector $(a_e)_{e\in E} \in \{-1,1\}^m$, such that $\sum_{e\in E} a_e x_e \equiv \beta \pmod p$.

(c) For each p-boundary β there is a vector $(a_e)_{e\in E} \in \{0,1\}^m$ such that $\sum_{e\in E} 2a_e x_e \equiv \beta \pmod p$.

Clearly, (a) is equivalent to (b). We now claim that (b) is equivalent to (c). To see this, simply do the following for each arc $\vec{e} = (v_i, v_j)$ of $\vec{G}$: add $f(e)$ to the j^{th}-entry of x_e and to $\beta(v_j)$, and subtract $f(e)$ from the i^{th}-entry of x_e and from $\beta(v_i)$. To deduce (c) from Corollary 5.8, what is left is to show that $\{a_e : e \in E\}$ can be decomposed into sufficiently many linear bases of $(\mathbb{Z}_p^n)_0$. This follows from the fact that G is $(8(p-1)(3p-4)+2p-4)$-edge-connected (and therefore contains $4(p-1)(3p-4)+p-2$ edge-disjoint spanning trees) and that the set of vectors a_e corresponding to the edges of a spanning tree of G forms a linear basis of $(\mathbb{Z}_p^n)_0$ (see [JLPT92]). □

A second proof consists in mimicking the proof of Theorem 5.6 (it turns out to give a better bound for the edge-connectivity of G).

Second proof of Theorem 5.3. As before, all values and operations are considered modulo p. We can assume without loss of generality that $f(E) \in \{1,2,\ldots,\frac{p-1}{2}\}$, since otherwise we can replace the value $f(e)$ of an edge e by $-f(e)$, without changing the problem.

We prove the result by induction on the number of vertices of G. The result is trivial if G contains only one vertex, so assume that G has at least two vertices.

For any $1 \leqslant i \leqslant k$, let E_i be the set of edges $e \in E$ with $f(e) = i$, and let $G_i = (V, E_i)$. Since G is $(6p-8)(p-1)$-edge-connected, G has minimum degree at least $(6p-8)(p-1)$ and then average degree at least $(6p-8)(p-1)$. As a consequence, there exists i such that G_i has average degree at least $12p-16$. By Lemma 5.13, since $\frac{12p-16}{4}+1 = 3p-3$, G_i has an induced subgraph $H = (X, F)$ with at least two vertices such that H is $(3p-3)$-edge-connected. Let G/X be the graph obtained from G by contracting X into a single vertex x (and removing possible loops). Since H contains more than one vertex, G/X has less vertices than G (note that possibly, $X = V$ and in this case G/X consists of the single vertex x). Since G is $(6p-8)(p-1)$-edge-connected, G/X is also $(6p-8)(p-1)$-edge-connected. Hence by the induction hypothesis it has an f-weighted β-orientation, where we consider the restriction of f to the edge-set of G/X, and we define $\beta(x) = \beta(X)$. Note that this orientation corresponds to an orientation of all the edges of G with at most one endpoint in X.

We now orient arbitrarily the edges of $G[X]$ not in F (the edge-set of H), and update the values of the p-boundary β accordingly (i.e. for each $v \in X$, we subtract from $\beta(v)$ the contribution of the arcs that were already oriented). It is easy to see that as the original β was a boundary, the new β is indeed a boundary. Finally, since all the edges of H have the same weight, and since H is $(3p-3)$-edge-connected, it follows from Corollary 5.11 that H has an f-weighted β-orientation (with respect to the updated boundary β). The orientations combine into an f-weighted β-orientation of G, as desired. □

Part II

Forbidding Patterns in Large Graphs

Chapter 6

From Local to Global

6.1 The Erdős-Hajnal conjecture

We now shift our attention to the behavior of large graphs not containing a given pattern. Fix a pattern H, recall that a graph G is *H-free* if it does not contain H as an induced subgraph. Part II will revolve around the following question: *Which common properties do all H-free graphs share?* One of the most well-known conjecture on this topic is due to Erdős and Hajnal in 1989 [EH89], which we discussed briefly in the Introduction, asserting that every H-free graph contains a large set of vertices either pairwise adjacent (i.e. a *clique*) or pairwise non-adjacent (i.e a *stable set*).

DEFINITION 6.1

> *We say that a pattern H has the Erdős-Hajnal property if there exists a constant $\varepsilon(H) > 0$ such that every H-free graph G contains either a clique or a stable set of size at least $|V(G)|^{\varepsilon(H)}$.*

Roughly speaking, a pattern H has the Erdős-Hajnal property if every *H-free* graph G contains a large (polynomially proportional to G) clique or stable set. It is worth noting [Erd47] that for almost every graph G, its largest clique and stable set are of size logarithmically proportional to G, i.e. $O(\log |V(G)|)$. In other words, if H has the Erdős-Hajnal property then all H-free graphs have an unusual structure in the sense that their maximum clique or stable set has size exponentially larger than expected.

The content of this chapter is covered in paper *Coloring tournaments: from local to global*, Ararat Harutyunyan, Tien-Nam Le, Stéphan Thomassé, and Hehui Wu, arXiv:1702.01607, and in a part of paper *Coloring dense digraphs*, Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé, *Combinatorica*, to appear.

The Erdős and Hajnal conjecture [EH89], however, asserts that every pattern has the Erdős-Hajnal property.

CONJECTURE 6.2 The Erdős-Hajnal conjecture [EH89]

Every graph has the Erdős-Hajnal property.

Very few *primitive*[1] patterns are known to have the Erdős-Hajnal property [Chu14], including C_5 – the cycle of length 5, P_4 – the path of length 4, the complement of P_4, and the *bull graph* (see Figure 6.1).

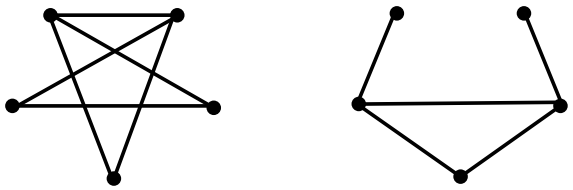

FIGURE 6.1 – Left: the complement of P_4, and right: the bull graph.

The Erdős-Hajnal conjecture – tournament version

Recall that a directed graph T is a *tournament* if there is exactly one arc between every pair of distinct vertices of T, and a tournament is *transitive* (or *acyclic*) it it does not contain any directed cycle. It is well-known that undirected graphs and tournaments are "equivalent" notions in the sense that every *enumerated* (i.e. vertices are enumerated) graph can be bijectively mapped to an enumerated tournament, where an edge is mapped to a forward arc and a non-edge is mapped to a backward arc. This mapping also shows that a transitive tournament is equivalent to either a clique or a stable set (depending on the direction of the arcs).

An equivalent version of the Erdős-Hajnal conjecture on tournament was proposed by Alon, Pach and Solymosi [APS01], where undirected graphs are replaced by tournaments and cliques and stables sets are replaced by transitive tournaments. For the rest of Part II, *the notion patterns will be used to mention tournaments instead of undirected graphs*. If a tournament T does not contain any subtournament isomorphic to a pattern H, we say that T is *H-free* (the term *induced subtournament* is unnecessary since a tournament contains exactly one arc between every pair of vertices). We start with the tournament version of the Erdős-Hajnal property.

DEFINITION 6.3

[1] A pattern created by "substituting" two patterns with the Erdős-Hajnal property also has the Erdős-Hajnal property; these pattern are not considered primitive. The definition of substitution of two graphs is beyond the scope of this thesis.

> *We say that a pattern H has the* Erdős-Hajnal property *if there exists a constant $\varepsilon(H) > 0$ such that every H-free tournament T contains a transitive tournament of size at least $|V(T)|^{\varepsilon(H)}$.*

And similarly, we have the tournament version of the Erdős-Hajnal conjecture.

CONJECTURE 6.4 The Erdős-Hajnal conjecture – tournament version, [APS01]

> *Every tournament has the Erdős-Hajnal property.*

The tournament version of the conjecture offers certain advantages comparing to its undirected version. For instance, given two undirected graphs H and H' where $H \subseteq H'$, it is possible that a graph G is H-free but not H'-free (for example, $P_2 \subseteq K_3$, and K_4 is P_2 free but not K_3-free). However, given two tournaments H and H' where $H \subseteq H'$, if a tournament T is H-free, then T is H'-free. This immediately implies that *if a tournament H has the Erdős-Hajnal property, then so does every subtournament of H*, which does not hold for the undirected version.

6.2 Celebrities and heroes

Similar to the undirected version, finding patterns satisfying the Erdős-Hajnal property for tournament is challenging. One may wish to address an extremal case of the conjecture: *Which patterns have the Erdős-Hajnal property where the largest transitive subtournament of T have linear size, i.e. the constant ε is 1?*

DEFINITION 6.5

> *We say that a pattern H is a* celebrity *if there exists a constant $\delta(H) > 0$ such that every H-free tournament T contains a transitive tournament of size at least $\delta(H)|V(T)|$.*

Similarly, *if H is a celebrity, then every subtournament of H is a celebrity*. A major result was achieved in 2013 [BCC+13], in which they completely characterized the family of celebrities and so added an infinite family of tournaments to the list of patterns satisfying the Erdős-Hajnal property. The proof of their result was not done via directly characterizing the family of celebrities, but another the family of patterns instead – the *heroes*. Unlike the celebrities, which forbidding forces the tournament to have a large transitive subtournament, forbidding a hero causes the tournament to be less colorful. We recall that a *vertex-coloring* (or *coloring* for short) of a digraph D is an assignment of each vertex of D to a color from a set of given colors. The definition of proper coloring digraph is slightly different from that of simple graphs: a coloring is *proper* if D does not contain any directed cycle of where all vertices have same color.

DEFINITION 6.6

We say that a pattern H is a hero if there exists a constant $c(H) > 0$ such that every H-free tournament T can be properly colored by $c(H)$ colors.

Coloring tournaments can be viewed alternatively as partitioning tournaments into subtournaments. A *partition* of a tournament T is a set of subtournaments $T_1, ..., T_k$ such that $\bigcup_{i=1}^{k} V(T_i) = V(T)$ and $V(T_i) \cap V(T_j) = \emptyset$ for every $i \neq j$.

PROPOSITION 6.7

A tournament can be properly colored by k colors if and only if it can be partitioned into k transitive subtournaments.

The proof is simple. If a tournament T can be colored by k colors, then consider a proper coloring of G with colors $1, ..., i$. For each color i, we assign T_i to be the subtournament of T on the set of vertices with color i. By the definition of proper coloring, each subtournament T_i is transitive. The converse direction is similar. It is now straightforward that *the family of heroes is contained in the family of celebrities*. Indeed, for every hero H, there is a constant $c(H)$ such that every H-free tournament T can be partitioned into $c(H)$ transitive tournaments. Then one of these tournament must have size at least $\frac{1}{c(H)}|V(T)|$. Hence H is a celebrity.

In the proof of Berger et al [BCC+13], they first showed that the two familys are identical.

THEOREM 6.8 [BCC+13]

Every celebrity is a hero.

Then they characterized all heroes as follows. Given tournaments H_1, H_2, H_3, we denote by $H_1 \Rightarrow H_2$ the vertex-disjoint union of H_1 and H_2 with all arcs from H_1 to H_2, and $\Delta(H_1, H_2, H_3)$ the vertex-disjoint union of H_1, H_2, H_3 with all arcs from H_1 to H_2, from H_2 to H_3, and from H_3 to H_1. For every integer $k \geqslant 1$, let T_k denote the transitive tournament on k vertices.

FIGURE 6.2 – Left: the tournament $C_3 \Rightarrow C_3$, and right: the tournament $\Delta(C_3, T_2, T_1)$.

THEOREM 6.9 [BCC+13]

The family of heroes are constructed as follows:

- *The singleton T_1 is a hero.*
- *If H_1 and H_2 are heroes, then $H_1 \Rightarrow H_2$ is a hero.*
- *If H is a hero, then $\Delta(H, T_k, T_1)$ and $\Delta(H, T_1, T_k)$ are heroes for every $k \geqslant 1$.*

Any tournament that cannot be constructed by this process is not a hero.

The combination of Theorems 6.8 with 6.9 gives a complete characterization of the set of celebrities, i.e. a complete characterization of patterns satisfying the Erdős-Hajnal property with constant $\varepsilon = 1$. A similar result was achieved one year later for the case $\varepsilon \geqslant 5/6$ [CCS14].

A conjecture by Berger et al

The main objective of Theorem 6.9 is to show that every H-free tournament T can be colored by a few colors. The authors observed that in order to reach that conclusion, they went cross in the proof multiple times the fact that every local region (the out-neighborhood a vertex) of T can be colored by a few colors as well. This leads to the question whether the local colorfulness of a tournament dictates its global colorfulness. In other words, whether it is true that if every local region of a tournament can be colored by a few colors, then so does the whole tournament. This is another vivid example about how local properties may affect the global behavior and structure of a large graph. More formally, given $t \geqslant 1$, a tournament T is *t-local* if for every vertex v, the subtournament of T on the set of out-neighbors of v can be colored by t colors.

CONJECTURE 6.10 Berger et al, 2013

There is a function f such that every t-local tournament T can be properly colored by $f(t)$ colors.

Conjecture 6.10 was proved recently for the case $t = 2$ in [CKL+18]. In this chapter, we will verify the conjecture for all t. We also generalize it to dense digraphs, which is a key result to extend the family of celebrities and the Erdős-Hajnal conjecture to dense digraphs in the next chapter.

6.3 Proof of Berger et al's conjecture

To proceed the proof of Conjecture 6.10, we need several definitions. We will provide the definitions in general digraphs for later use. The *chromatic number* $\chi(D)$ of a di-

graph D, is the minimum number k such that D can be properly colored by k colors. Given a subset A of $V(D)$, we usually denote by $\chi(A)$ the chromatic number of the digraph $D[A]$ when it is clear from the context.

Given a digraph D and a set $S \subset V(D)$, we say that S is a *dominating set* of D if every vertex in $V(D) \setminus S$ has an in-neighbor in S. The *domination number* $\gamma(D)$ of a digraph D is the smallest number k such that D has a dominating set of size k. Recall that for every vertex v in a digraph D, we denote by $N_D^+(v)$ the set of out-neighbors of v in D. Given a subset X of $V(D)$, let $N_D^+(X)$ denote the union of all $N_D^+(v)$, for $v \in X$, and denote by $N_D^+[X] := X \cup N_D^+(X)$.

We first have the following property of t-local tournaments.

PROPOSITION 6.11

Let T be a t-local tournament. Then for every subset X of $V(T)$, $\chi(N^+[X]) \leqslant t|X|$.

Proof. Observe that for every v, we can color $N^+[v]$ with t colors by first coloring $N^+(v)$ with t colors and then coloring arbitrarily v with one of these t colors (since v dominates $N^+(v)$, coloring v does not create any monochromatic directed cycle). To color $N^+[X]$, we color $N^+[v]$ sequentially for every $v \in X$. For each $v \in X$, we color the uncolored vertices in $N^+[v]$ with t new colors. Thus we can color $N^+[X]$ with $t|X|$ colors. □

Let $\mathcal{C}$ be a class of tournaments closed under taking subtournaments. We say that $\mathcal{C}$ is *tamed* if for every integer k there exists K and ℓ such that every tournament $T \in \mathcal{C}$ with $\chi(T) \geqslant K$ contains a set A of ℓ vertices such that $\chi(A) \geqslant k$. Note that a class of tournaments with bounded chromatic number is indeed tamed.

THEOREM 6.12

For every t, the class of t-local tournaments is tamed.

Thank to Theorem 6.12, we can show that Conjecture 6.10 is true.

THEOREM 6.13

There is a function f such that every t-local tournament T satisfies $\chi(T) \leqslant f(t)$.

Proof. Since the class of t-local tournaments is tamed, by the tamed property applied to $k = t + 1$, there exists (K, ℓ) such that every t-local tournament T with $\chi(T) \geqslant K$ contains a set A of ℓ vertices and $\chi(A) \geqslant t + 1$. If a vertex $v \in V(T) \setminus A$ does not have an in-neighbor in A, then $A \subseteq N^+(v)$, and so $t + 1 \leqslant \chi(A) \leqslant \chi(N^+(v)) \leqslant t$, a contradiction. Hence, A is a dominating set of T. Thus, $\chi(T) = \chi(N^+[A]) \leqslant t|A| = \ell t$. Consequently, t-local tournaments have chromatic number at most $f(t) := \max(K, \ell t)$. □

We now prove Theorem 6.12.

Proof of Theorem 6.12. We fix some arbitrary t and show the property by induction on k. The claim is trivial for $k = 1$. For $k = 2$, we can choose $K = 2$ and $\ell = 3$. Indeed, if a tournament T satisfies $\chi(T) \geqslant K = 2$, it contains an oriented triangle S of size $\ell = 3$ and $\chi(S) \geqslant k = 2$.

Assuming now that (K, ℓ) exists for k, we want to find (K', ℓ') for $k+1$. For this, we set $K' := 2kt(K + \ell t + 1)$, and fix ℓ' later. Let T be a t-local tournament such that $\chi(T) \geqslant K'$. Let D be a dominating set of T of minimum size d. By Proposition 6.11, we have $\chi(T) = \chi(N^+[D]) \leqslant td$. In particular, $d \geqslant 2k(K + \ell t + 1)$. Consider a subset W of D of size $k(K + \ell t + 1)$. By Proposition 6.11, we have $\chi(N^+[W]) \leqslant kt(K + \ell t + 1)$, and hence

$$\chi(V \setminus N^+[W]) \geqslant \chi(T) - \chi(N^+[W]) \geqslant kt(K + \ell t + 1) \geqslant K,$$

where V is the vertex set of T. In particular, by the tamed property applied to k, one can find a set $A \subseteq V \setminus N^+[W]$ such that A has ℓ vertices and $\chi(A) \geqslant k$. Note that by construction, $A \cap W = \emptyset$ and all arcs between A and W are directed from A to W.

Consider now a subset S of W of size $K + \ell t + 1$. We claim that $\chi(N^+(S)) \geqslant K + \ell t$. If not, we can cover $N^+(S)$ by at most $K + \ell t - 1$ transitive sets. Since every transitive tournament has a dominating set of size 1, we can choose a dominating set S' of $N^+(S)$ of size at most $K + \ell t - 1$. Note that x dominates S for any $x \in A$, and so $S' \cup \{x\}$ dominates $N^+[S]$. Hence $(D \setminus S) \cup S' \cup \{x\}$ would be a dominating set of T of size less than $|D|$, which contradicts the minimality of $|D|$. Therefore, $\chi(N^+(S)) \geqslant K + \ell t$.

Let N' be the set of vertices $N^+(S) \setminus N^+(A)$. Observe that all the arcs between N' and A are directed from N' to A, and that by Proposition 6.11 applied to $N^+(A)$, we have $\chi(N^+(A)) \leqslant \ell t$, and so

$$\chi(N') \geqslant \chi(N^+(S)) - \chi(N^+(A)) \geqslant K + \ell t - \ell t = K.$$

Thus, by the tamed property applied to k, there is a subset A_S of N' such that $|A_S| = \ell$ and $\chi(A_S) \geqslant k$. Note that by construction, $A_S \cap A = \emptyset$ and all arcs between A_S and A are directed from A_S to A.

We now construct our subtournament of T with chromatic number at least $k+1$. For this we consider the set of vertices $A \cup W$ to which we add the collection of A_S, for all subsets $S \subseteq W$ of size $K + \ell t + 1$. Call A' this new tournament and observe that its number of vertices is at most

$$\ell' := \ell + k(K + \ell t + 1) + \ell \binom{k(K + \ell t + 1)}{K + \ell t + 1}.$$

To conclude, it is sufficient to show that $\chi(A') \geqslant k + 1$. Suppose not, and for contradiction, take a k-coloring of A'. Since $|W| = k(K + \ell t + 1)$ there is a monochromatic

set S in W of size $K + \ell t + 1$ (say, colored 1). Recall that we have all arcs from A_S to A and all arcs from A to S, and note that since $\chi(A) \geqslant k$ and $\chi(A_S) \geqslant k$, both A and A_S have a vertex of each of the k colors. Hence there are $u \in A$ and $w \in A_S$ colored 1. Since $A_S \subseteq N^+(S)$, there is $v \in S$ dominating w. We then obtain the monochromatic cycle uvw of color 1, a contradiction. Thus, $\chi(A') \geqslant k + 1$, completing the proof. □

6.4 Generalizing to dense digraphs

The term digraphs in Part II is used to mention *simple, loopless and finite digraphs*, where a digraph D is *simple* if for every two vertices u and v of D, there is at most one arc with endpoints $\{u, v\}$. The purpose of this section is to extending Conjecture 6.10 from tournaments to *dense digraphs*, which will serve as a key tool for the results in the next chapter.

Besides the out- and in-neighbors $N_D^-(v)$ and $N_D^+(v)$ of a vertex v in a digraph D, we also denote by $N_D^o(v)$ the set of vertices non-adjacent to v in D. For a subset X of $V(D)$, we denote by $N_D^o(X)$ the set of vertices of V non-adjacent to at least one vertex in X. A *stable set* in a digraph is a set of vertices pairwise non-adjacent. Given a positive integer α, a digraph D is called *α-dense*[2] if it contains no stable set of size $\alpha + 1$. Given $t \geqslant 1$, a digraph D is *t-local* if for every vertex v we have $\chi(N^+(v)) \leqslant t$. The extension of Conjecture 6.10 to dense digraphs is as follows.

THEOREM 6.14

For every positive integers α and t, there is a function $f_\alpha(t)$ such that every α-dense t-local digraph D can be properly colored by $f_\alpha(t)$ colors.

And we have a similar version of Proposition 6.11.

PROPOSITION 6.15

Given a t-local digraph D, for every subset X of $V(D)$, $\chi\big(N^+[X]\big) \leqslant t|X|$.

We also have another observation regarding dominating sets in an acyclic digraph.

PROPOSITION 6.16

An acyclic digraph D has a dominating set which is also a stable set.

Proof. We proceed by induction on $|D|$ to show that every acyclic digraph D has a dominating set S which is stable. The statement clearly holds for $|D| = 1$. For $|D| > 1$, since D is acyclic, there is a vertex v with no in-neighbors. Then $V(D)\backslash\{v\} =$

[2] In liturature, an α-dense digraph is called a digraph with *independence number* at most α.

$N^+(v) \cup N^o(v)$. Applying induction to $D[N^o(v)]$, we obtain a stable dominating set S' of $D[N^o(v)]$. Clearly $S := S' \cup \{v\}$ is a dominating set of D. Note that $S' \subseteq N^o(v)$, and so S is stable. □

Let $\mathcal{C}$ be a class of digraphs closed under taking subdigraphs. We say that $\mathcal{C}$ is *tamed* if for every integer k there exists K and ℓ such that every digraph $T \in \mathcal{C}$ with $\chi(T) \geqslant K$ contains a set A of ℓ vertices such that $\chi(A) \geqslant k$. Note that a class of digraphs with bounded chromatic number is indeed tamed.

We proceed the proof of Theorem 6.14 by induction on α. The case $\alpha = 1$ is Conjecture 6.10. The following lemma is analogous to Theorem 6.12.

LEMMA 6.17

Suppose that Theorem 6.14 is true for $\alpha - 1$, i.e., $f_{\alpha-1}(t)$ exists for every t. For every t, the class of α-dense t-local digraphs is tamed.

We now can finish the proof of Theorem 6.14.

Proof of Theorem 6.14. Suppose that the theorem is true for $\alpha - 1$, i.e., $f_{\alpha-1}(t)$ exists for every t. We now show that it is true for α. Since the class of α-dense t-local digraphs is tamed, by applying tamed property for $k = t + f_{\alpha-1}(t) + 1$, we have that there exist (K, ℓ) such that every α-dense t-local digraph D $\chi(D) \geqslant K$ contains a set A of ℓ vertices and $\chi(A) \geqslant t + f_{\alpha-1}(t) + 1$. We claim that A is a dominating set. If not, then there is a vertex v such that $A \subseteq N^o(v) \cup N^+(v)$. Then $\chi(A) \leqslant \chi(N^o(v)) + \chi(N^+(v)) \leqslant t + f_{\alpha-1}(t)$, a contradiction. Hence, A is a dominating set of D. Thus, $\chi(D) = \chi(\bigcup_{x\in A}(N^+(x) \cup \{x\})) \leqslant (t+1)|A| = \ell + \ell t$. Consequently, t-local digraphs have chromatic number at most $f(t) := \max(K, \ell + \ell t)$. □

Proof. We fix some arbitrary t and show the property by induction on k. The claim is trivial for $k = 1$. Assuming now that (K, ℓ) exists for k, we want to find (K', ℓ') for $k+1$. For this, we set $s := K + \ell f_{\alpha-1}(t) + \ell t$ and

$$K' := 2k(\alpha s + 1)(t + f_{\alpha-1}(t) + 1),$$

and fix ℓ' later.

Let D be an α-dense t-local digraph with vertex set V such that $\chi(D) \geqslant K'$. Let B be a dominating set of D of minimum size b. By Proposition 6.15, we have $\chi(D) = \chi(N^+[B]) \leqslant tb$. In particular, $b \geqslant K'/t \geqslant 2k(\alpha s + 1)$. Consider a subset W of B of size $k(\alpha s + 1)$. By Proposition 6.15, we have $\chi(N^+(W)) \leqslant kt(\alpha s + 1)$. By induction hypothesis on $\alpha - 1$, for every $x \in W$, we have $\chi(N^o(x)) \leqslant f_{\alpha-1}(t)$ since $N^o(x)$ is $(\alpha - 1)$-dense and t-local.

Let $M^-(X) := \bigcap_{x\in X} N^-(x)$ for any $X \subseteq D$, the set of common in-neighbors of all vertices of X. Then we have,

$$\begin{aligned}\chi(M^-(W)) &\geqslant \chi(D) - \chi(N^+[W]) - \chi(N^o(W))\\ &\geqslant K' - kt(\alpha s+1) - \sum_{x\in W}\chi(N^o(x))\\ &\geqslant K' - k(\alpha s+1)(t+f_{\alpha-1}(t)+1)\\ &\geqslant K'/2\\ &\geqslant K.\end{aligned}$$

In particular, by the tamed property applied to k, one can find a set $A \subset M^-(W)$ such that A has ℓ vertices and $\chi(A) \geqslant k$.

Consider now a subset S of W of size $\alpha s+1$. We claim that $\chi(N^+(S)) \geqslant s$. If not, we can cover $N^+(S)$ by at most $s-1$ acyclic sets. Since every acyclic set has independence number at most α, it has a dominating set of size at most α by Proposition 6.16. Hence $N^+(S)$ has a dominating set, say S' of size at most $\alpha(s-1) \leqslant |S|-2$. But this yields a contradiction since the set $(B\setminus S)\cup S'\cup\{x\}$, where x is an arbitrary vertex in A, would be a dominating set of T of size less than $|B|$. Therefore, $\chi(N^+(S)) \geqslant s$.

By Proposition 6.15 applied to $N^+(A)$, we have $\chi(N^+(A)) \leqslant \ell t$. Hence

$$\begin{aligned}\chi(N^+(S)\cap M^-(A)) &\geqslant \chi(N^+(S)) - \chi(N^+(A)) - \chi(N^o(A)),\\ &\geqslant s - \ell t - \sum_{x\in A}\chi(N^o(x)),\\ &\geqslant (K+\ell f_{\alpha-1}(t)+\ell t) - \ell t - |A| f_{\alpha-1}(t),\\ &= K.\end{aligned}$$

Thus, by the tamed property applied to k, there is a subset A_S of $N^+(S)\cap M^-(A)$ such that $|A_S| = \ell$ and $\chi(A_S) \geqslant k$.

We now construct our subset of V with chromatic number at least $k+1$. For this we consider the set A' consisting of vertices $A\cup W$ to which we add the collection of A_S, for all subsets $S\subseteq W$ of size $\alpha s+1$. Observe that the number of vertices of A' is at most

$$\ell' := \ell + k(\alpha s+1) + \ell k(\alpha s+1)\binom{s}{\alpha} + 1.$$

To conclude, it is sufficient to show that $\chi(A') \geqslant k+1$. Suppose not, and for contradiction, take a k-coloring of A'. Since $|W| = k(\alpha s+1)$ there is a monochromatic set S in W of size $\alpha s+1$ (say, colored 1). Recall that $A_S \subseteq M^-(A)$ and $A \subseteq M^-(W) \subseteq M^-(S)$, so we have all arcs from A_S to A and all arcs from A to S, and note that since $\chi(A) \geqslant k$ and $\chi(A_S) \geqslant k$, both A and A_S have a vertex of each of the k colors. Hence there are $u \in A$ and $w \in A_S$ colored 1. Since $A_S \subseteq N^+(S)$, there is $v \in S$ such that vw is an arc. We then obtain the monochromatic cycle uvw of color 1, a contradiction. Thus, $\chi(A') \geqslant k+1$, completing the proof of the claim. □

We conclude the chapter by mentioning another version of Theorem 6.14, where the tamed property is applied to in-neighbors instead of out-neighbors. The theorem can be obtained easily form Theorem 6.14 by just reversing the directions of all arcs.

THEOREM 6.18

For every t and α, there is $f_\alpha(t)$ such that every α-dense digraph D with $\chi(N^-(v)) \leqslant t$ for every $v \in V(D)$ has chromatic number at most $f_\alpha(t)$.

Chapter 7

Generalizing the Erdős-Hajnal conjecture

7.1 The Erdős-Hajnal property in digraphs

Recall the Erdős-Hajnal conjecture that *for every tournament H, there exists a constant $\varepsilon(H) > 0$ such that every H-free tournament T contains a transitive tournament with at least $|V(T)|^{\varepsilon(H)}$ vertices.* Since the conjecture is confined in tournaments, it is natural to ask whether the result can be extended to general digraphs. This was in fact a conjecture by Harutyunyan and McDiarmid [HM12]. A set of vertices A in a digraph D is *acyclic* if the induced subgraph $D[A]$ is acyclic.

DEFINITION 7.1

> *We say that a digraph H has the general Erdős-Hajnal property if there exists a constant $\varepsilon(H) > 0$ such that every H-free digraph D contains an acyclic set of size at least $|V(D)|^{\varepsilon(H)}$.*

Harutyunyan and McDiarmid [HM12] conjectured that every digraph has the general Erdős-Hajnal property.

CONJECTURE 7.2 [HM12]

> *Every digraph has the general Erdős-Hajnal property.*

The content of this chapter is covered in paper *Coloring dense digraphs*, Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé, *Combinatorica*, to appear.

Conjecture 7.2 should be substantially harder than the Erdős-Hajnal conjecture (since it generalizes the class of both patterns and large graphs from tournaments to digraphs). Even a very simple pattern as the directed triangle C_3 is not known to have the general Erdős-Hajnal property. So far, the transitive tournaments T_k are the only digraphs known to have the general property. The proof is quite simple. Let D be a T_k-free digraph. We enumerate the vertices of D arbitrarily and consider a mapping from D to an undirected graph G where a forward arc of D is mapped to an edge of G, and a backward arc or a non-edge of D is mapped to a non-arc of G. Then G does not have any clique of size k, and so has a stable set S of linear size. Note that the induced subgraph $D[S]$ contains only backward arcs or non-edges. Since the underlying graph of $D[S]$ contains no clique of size k, it must contain a stable set of linear size, which is also a stable set in D.

While we could not provide any non-trivial pattern having the general Erdős-Hajnal property, we will show in this chapter a family of patterns having a weaker property than the general one (which requires to hold for all H-free digraphs) but stronger than the original one (requiring to hold for all H-free tournaments), which we called the *extended Erdős-Hajnal property* (requiring to hold for all H-free dense digraphs). Recall that given a positive integer α, a digraph D is called *α-dense* if it contains no stable set of size $\alpha + 1$.

DEFINITION 7.3

> *We say that a digraph H has the* extended Erdős-Hajnal property *if there exists a constant $\varepsilon(H, \alpha) > 0$ such that every H-free digraph D contains an acyclic subdigraph of size at least $|V(D)|^{\varepsilon(H,\alpha)}$.*

Clearly, every tournament with the general Erdős-Hajnal property must have the extended one, and every tournament with the extended property must have the original one. One idea to prove Conjecture 7.2, which asserts that every digraph has the general property, is to show that the converse of the two observations above holds as well, i.e., every tournament with the original property has the extended one, and every tournament with the extended property has the general one. To support Conjecture 7.2, we will prove in this chapter that the family of celebrities have the extended Erdős-Hajnal property, and so characterize all tournaments having the extended property with a linear size acyclic set.

DEFINITION 7.4

> *We say that a digraph H is a* superstar *if for every positive integer α, there exists a constant $\delta(H, \alpha) > 0$ such that every α-dense H-free digraph D contains an acyclic subgraph of size at least $\delta(H, \alpha)|V(D)|$.*

The first observation is that *every superstar is a celebrity*. In the rest of this chapter, we will show that the converse is true.

THEOREM 7.5

Every superstar is a celebrity.

To prove Theorem 7.5, we follow Berger et al [BCC+13] and resort to the coloring of digraphs.

DEFINITION 7.6

We say that a digraph H is a superhero *if for every positive integer α, there exists a constant $c(H,\alpha) > 0$ such that every α-dense H-free digraph D can be properly colored by $c(H,\alpha)$ colors.*

Similarly, it is not hard to see that every superhero is a superstar. Thus, we have:

$$\mathcal{F}_{\text{superheroes}} \subseteq \mathcal{F}_{\text{superstars}} \subseteq \mathcal{F}_{\text{heroes}} = \mathcal{F}_{\text{celebrity}},$$

where $\mathcal{F}_x$ is the family of patterns x. Let us recall that given tournaments H_1, H_2, H_3, we denote by $H_1 \Rightarrow H_2$ the vertex-disjoint union of H_1 and H_2 with complete arcs from H_1 to H_2, and $\Delta(H_1, H_2, H_3)$ the vertex-disjoint union of H_1, H_2, H_3 with complete arcs from H_1 to H_2, from H_2 to H_3, and from H_3 to H_1. For every integer $k \geqslant 1$, let T_k denote the transitive tournament on k vertices. To prove Theorem 7.5, it is sufficient to show that every hero is a superhero. More precisely, it is sufficient to prove the following theorem, which was a conjecture of Aboulker, Charbit and Naserasr (private communication, 2016).

THEOREM 7.7

The family of superheroes are constructed as follows:

- *The singleton T_1 is a superhero.*
- *If H_1 and H_2 are superheroes, then $H_1 \Rightarrow H_2$ is a superhero.*
- *If H is a superhero, then $\Delta(H, T_k, T_1)$ and $\Delta(H, T_1, T_k)$ are superheroes for every $k \geqslant 1$.*

Any tournament that cannot be constructed by this process is not a superhero.

Obviously T_1 is a superhero. Thus to prove Theorem 7.7, it suffices to prove the following two theorems.

THEOREM 7.8

If H_1 and H_2 are superheroes, then $H_1 \Rightarrow H_2$ is a superhero.

THEOREM 7.9

If H is a superhero, then $\Delta(H, T_1, T_k)$ and $\Delta(H, T_k, T_1)$ are superheroes for any $k \geqslant 1$.

Notation and remarks

The proofs of Theorems 7.8 and 7.9 are considerably technical. Thus, we recall here thoroughly the necessary notation. Given a digraph D, we say that u *sees* v and v is *seen by* u if uv is an arc in D. For every $v \in V(D)$, we denote by $N_D^+(v)$ (res. $N_D^-(v)$) the set of out-neighbors (res. in-neighbors) of v in D. Let $N(v) = N_D^+(v) \cup N_D^-(v)$. For every $X \subseteq V(D)$, let $N_D^+(X) = \bigcup_{v \in X} N_D^+(v)$ (res. $N_D^-(X) = \bigcup_{v \in X} N_D^-(v)$), the set of vertices seen by (res. seeing) at least one vertex of X, and let $M_D^+(X) = \bigcap_{v \in X} N_D^+(v)$ (res. $M_D^-(X) = \bigcap_{v \in X} N_D^-(v)$), the set of vertices seen by (res. seeing) all vertices of X. Let $N_D(X)$ denote $N_D^+(X) \cup N_D^-(X)$.

For every $v \in V(D)$, we denote by $N_D^o(v)$ the set non-adjacent vertices of v in D. For a subset X of $V(D)$, we denote by $N_D^o(X)$ the set of vertices of V non-adjacent to at least one vertex of X. When it is clear in the context (most of the time), we omit the subscript D in this notation. We will use throughout this chapter the fact that $V(D) \setminus X = M_D^+(X) \cup N_D^-(X) \cup N_D^o(X)$ for any $X \subseteq V(D)$. Given a digraph D and two disjoint sets $X, Y \subseteq V(D)$, we denote $X \to_D Y$ (or just $X \to Y$) if there is no arc from Y to X in D. A key observation is that if $X \to Y$, then $\chi(X \cup Y) = \max\big(\chi(X), \chi(Y)\big)$.

A side remark is that some proofs in this chapter proceeding by induction on α use the fact that if D is α-dense, then $N^o(v)$ is $(\alpha - 1)$-dense for every v, and thus $\chi(N^o(v))$ is bounded. In these inductive proofs, we often cite known results on tournaments for the base case $\alpha = 1$. However, the proofs here are indeed self-contained since to prove the base case $\alpha = 1$, we just repeat the same arguments and use the fact that in a tournament, $N^o(v) = \emptyset$ for any vertex v. Hence for example, the proof of Theorem 7.9 can serve as an alternative proof for Theorem 4.1 in [BCC+13] (that if H is a hero, then so are $\Delta(H, T_k, T_1)$ and $\Delta(H, T_1, T_k)$).

7.2 Chains of superheroes

In this section, we will prove Theorem 7.8, which states that if H_1 and H_2 are superheroes, then so is $H_1 \Rightarrow H_2$. We will reuse the notions of *r-mountains* and *(r, s)-cliques* introduced in [BCC+13]. Let us first give the idea of the proof of Theorem 7.8 for a special case: $(C_3 \Rightarrow C_3)$-free tournaments have bounded chromatic number. Given a $(C_3 \Rightarrow C_3)$-free tournament T, suppose that there is a small set Q in T with chromatic number 3. Then for any partition of Q into Q_1, Q_2, at least one part of the partition has chromatic number at least 2, and so contains a copy of C_3. Let $Y_{Q_1,Q_2} \subseteq V(D) \setminus Q$ be the set of vertices seeing all vertices of Q_1 and seen by all vertices of Q_2. Observe

that Y_{Q_1,Q_2} is C_3-free, otherwise a copy of C_3 in Y_{Q_1,Q_2} together with a copy of C_3 in either Q_1 or Q_2 forms a copy of $C_3 \Rightarrow C_3$. Note that $V(T)\setminus Q$ is covered by only $2^{|Q|}$ such sets Y_{Q_1,Q_2}, and hence $\chi(T)$ is bounded.

Hence we wish to find such a small set of vertices Q with chromatic number 3. To this end, we call an arc uv of T *thick* if $N^-(u) \cap N^+(v)$ contains a copy of C_3. If T has no thick arcs, then intuitively T should have simple structure, and thus, bounded chromatic number. Suppose that T contains a (not necessarily directed) triangle uvw where all of the three arcs are thick. Then for each of the three thick arcs, we take its thickness-certificate (i.e., a copy of C_3) and together with u, v, w we obtain a set Q of at most 12 vertices. It is straightforward to verify that Q has chromatic number at least 3, and thus, by the argument above $\chi(T)$ is bounded. If T contains no triangle of thick arcs, then for any vertex v, the set of vertices adjacent to v by a thick arc induces a thick-arc-free tournament, which, intuitively, should have bounded chromatic number. We then easily bound the chromatic number of the sets of non-thick in-neighbors and non-thick out-neighbors of v, and hence bound the chromatic number of T.

The proof of the general case is in the same vein. Intuitively, we search for a small set Q with large chromatic number as described above. We will capture the notion of the set Q with the definition of an object called an *r-mountain*, and the notion of a triangle of thick arcs with objects called *(r,s)-cliques*. Given a digraph D, the formal definitions (which are borrowed from [BCC+13]) of *r-thick-arc*, *(r,s)-clique*, and *r-mountain* in D are defined inductively on r as follows. Every vertex of D is a 1-mountain. For every $r, s \geqslant 1$,

- An arc e of D is *r-thick* if $N^-(u) \cap N^+(v)$ contains an r-mountain. An r-mountain in $N^-(u) \cap N^+(v)$ is a *certificate* of r-thickness of e.
- An *(r,s)-clique* of D is a set $S \subseteq V(D)$ such that $|S| = s$, and for every distinct vertices $u, v \in S$, either uv or vu is an arc that is r-thick.
- Given an $(r, r+1)$-clique S and a certificate $C_{u,v}$ for every distinct $u, v \in S$, then the tournament induced on $S \cup (\bigcup_{u,v\in S} C_{u,v})$ is an *$(r+1)$-mountain* of D.

Note that if a digraph D contains an $(r, r+1)$-clique, then D contains an $(r+1)$-mountain, which is the $(r, r+1)$-clique together with certificates of all r-thick arcs of that $(r, r+1)$-clique. Hence, if D contains no $(r+1)$-mountain, then D contains no $(r, r+1)$-clique.

LEMMA 7.10 [BCC+13], Lemma 3.3

> *Every r-mountain has chromatic number at least r, and has at most $(r!)^2$ vertices.*

Fix two superheroes H_1 and H_2.

LEMMA 7.11

Fix $\alpha \geqslant 2$ and $r, s \geqslant 1$, suppose that there are b_0, b_1, b_2 such that

- *Every $(\alpha-1)$-dense $(H_1 \Rightarrow H_2)$-free digraph D has $\chi(D) < b_0$.*
- *Every α-dense $(H_1 \Rightarrow H_2)$-free digraph D containing no (r,s)-clique has $\chi(D) < b_1$.*
- *Every α-dense $(H_1 \Rightarrow H_2)$-free digraph D containing no r-mountain has $\chi(D) < b_2$.*

Then there is b_3 such that every α-dense $(H_1 \Rightarrow H_2)$-free digraph D containing no $(r, s+1)$-clique has $\chi(D) < b_3$.

Proof. A small remark is that the second hypothesis seems redundant since if D contains no $(r,2)$-clique, then D contains no r-thick arc, and so contains no r-mountain. However, the second hypothesis is necessary for the case s=1.

First, note that since H_1 and H_2 are superheroes, there is b_4 such that every α-dense H_1-free (or H_2-free) digraph D has $\chi(D) \leqslant b_4$. We first identify all r-thick arcs of D. Fix an arbitrary vertex v. Then $V(D)\backslash\{v\}$ can be partitioned into four sets: N^* the set of neighbors of v that are connected to v by an r-thick arc; $N^- = N^-(v)\backslash N^*$; $N^+ = N^+(v)\backslash N^*$; and $N^o(v)$.

Note that $N^o(v)$ is $(H_1 \Rightarrow H_2)$-free and $(\alpha-1)$-dense, and so by the first hypothesis, $\chi(N^o(v)) \leqslant b_0$. The crucial fact is that the digraph induced by the set N^* does not contain an (r,s)-clique; indeed, an (r,s)-clique together with v would form an $(r, s+1)$-clique, a contradiction to the fact that D has no $(r, s+1)$-cliques. Hence by the second hypothesis, $\chi(N^*) \leqslant b_1$.

CLAIM 7.12

There is b_5 such that either $\chi(N^-) \leqslant b_5$ or $\chi(N^+) \leqslant b_5$.

Proof. Suppose that $\chi(N^-) \geqslant b_4$, then N^- contains a copy of H_1, say $\widehat{H}_1$. Note that

$$N^+ = \big(M^+(\widehat{H}_1) \cap N^+\big) \cup \big(N^o(\widehat{H}_1) \cap N^+\big) \cup \big(N^-(\widehat{H}_1) \cap N^+\big).$$

- If $\chi(M^+(\widehat{H}_1) \cap N^+) \geqslant b_4$, then $M^+(\widehat{H}_1) \cap N^+$ contains a copy of H_2, say $\widehat{H}_2$, and we have $\widehat{H}_1 \Rightarrow \widehat{H}_2$ forming a copy of $H_1 \Rightarrow H_2$, a contradiction. Hence, $\chi(M^+(\widehat{H}_1) \cap N^+) \leqslant b_4$
- For each $u \in \widehat{H}_1$, we have $N^o(u) \cap N^+$ is $(\alpha-1)$-dense, so $\chi(N^o(u) \cap N^+) \leqslant b_0$.
- For each $u \in \widehat{H}_1$, if $\chi(N^-(u) \cap N^+) \geqslant b_2$, then $N^-(u) \cap N^+$ contains a r-mountain. This means that uv is an r-thick arc, contradicting $u \notin N^*$. Hence $\chi(N^-(u) \cap N^+) \leqslant b_2$, for each $u \in \widehat{H}_1$.

Thus we have,

$$\begin{aligned}\chi(N^+) &\leqslant \chi(M^+(\widehat{H}_1) \cap N^+) + \sum_{u \in \widehat{H}_1} \Big(\chi(N^o(u) \cap N^+) + \chi(N^-(u) \cap N^+)\Big) \\ &\leqslant b_4 + |H_1|(b_0 + b_2).\end{aligned}$$

Set $b_5 := b_4 + |H_1|(b_0 + b_2)$. We have just shown that if $\chi(N^-) \geqslant b_4$, then $\chi(N^+) \leqslant b_5$. Hence either $\chi(N^-) \leqslant b_4 \leqslant b_5$ or $\chi(N^+) \leqslant b_5$. This proves the claim. □

Note that $N^+(v) \subseteq N^+ \cup N^*$, and so $\chi(N^+(v)) \leqslant \chi(N^+) + \chi(N^*)$. Similarly, $\chi(N^-(v)) \leqslant \chi(N^-) + \chi(N^*)$. Hence for every $v \in V(D)$, either $\chi(N^+(v)) \leqslant b_5 + b_1$ or $\chi(N^-(v)) \leqslant b_5 + b_1$. Let R be the set of all vertices $v \in V(D)$ with $\chi(N^+(v)) \leqslant b_5 + b_1$ and B be the set of all vertices $v \in V(D)$ with $\chi(N^-(v)) \leqslant b_5 + b_1$. Note that $R \cup B = V(D)$.

Observe that R is an α-dense digraph, and $\chi(N_R^+(v)) \leqslant \chi(N_D^+(v)) \leqslant b_5 + b_1$ for every $v \in R$. Then applying Theorem 6.14 to R with $t = b_5 + b_1$, there is b_6 such that $\chi(R) \leqslant b_6$. Similarly, by Theorem 6.18, there is b_7 such that $\chi(B) \leqslant b_7$. Hence $\chi(D) \leqslant \chi(R) + \chi(B) \leqslant b_6 + b_7$. Setting $b_3 := b_6 + b_7$ completes the proof of Lemma 7.11. □

Recall that if D contains no $(r+1)$-mountain, then D contains no $(r, r+1)$-clique. We are now ready to show that digraphs which do not contain a mountain have bounded chromatic number.

LEMMA 7.13

Let $\alpha \geqslant 2$, and suppose that every $(\alpha-1)$-dense $(H_1 \Rightarrow H_2)$-free digraph D has $\chi(D) < b_0$ for some b_0. Then for every r, there exists $g_\alpha(r)$ such that every α-dense $(H_1 \Rightarrow H_2)$-free digraph not containing an r-mountain has $\chi(D) \leqslant g_\alpha(r)$.

Proof. We proceed by induction on r. If D contains no 1-mountain, then D has no vertices, and we can set $g_\alpha(r) := 0$. Now suppose by induction that $g_\alpha(r)$ exists. We will show that $g_\alpha(r+1)$ exists. First, we claim the following.

(A) For every s, there exists function $g'_{\alpha,r}(s)$ such that if D contains no (r,s)-clique, then $\chi(D) \leqslant g'_{\alpha,r}(s)$.

We prove (A) by induction on s. For $s = 1$, if D contains no $(r,1)$-clique, then D has no vertex, so $g'_{\alpha,r}(s) = 0$. Suppose, by induction, that $g'_{\alpha,r}(s)$ exists. Let D be a digraph not containing a $(r, s+1)$-clique. Applying Lemma 7.11 with $b_1 = g'_{\alpha,r}(s)$ and $b_2 = g_\alpha(r)$, we deduce that $g'_{\alpha,r}(s+1)$ exists. This proves (A).

If D contains no $(r+1)$-mountain, then D contains no $(r, r+1)$-clique, implying $\chi(D) \leqslant g'_{\alpha,r}(r+1)$. Set $g_\alpha(r+1) := g'_{\alpha,r}(r+1)$. This completes the proof. □

To prove Theorem 7.8, it suffices to prove the following lemma.

LEMMA 7.14

For every integer $\alpha \geqslant 1$, there exists $f(\alpha)$ such that every α-dense $(H_1 \Rightarrow H_2)$-free digraph D has $\chi(D) \leqslant f(\alpha)$.

Proof. We proceed by induction on α. Since $H_1 \Rightarrow H_2$ is a hero (see [BCC+13], Theorem 3.2), Lemma 7.14 is true for $\alpha = 1$. Suppose that Lemma 7.14 is true for $\alpha - 1$ and let $c_0 = f(\alpha - 1)$. Since both H_1 and H_2 are superheroes, there exists c_1 such that if D is any H_1-free or H_2-free digraph, then $\chi(D) \leqslant c_1$. Let D be an α-dense $(H_1 \Rightarrow H_2)$-free digraph with vertex set V. If D does not contain a $(c_0 + 2c_1)$-mountain, then by applying Lemma 7.13 to D with $b_0 = c_0$, there is c_2 such that $\chi(D) \leqslant c_2$. Thus, it remains to consider the case that D contains a $(c_0 + 2c_1)$-mountain.

CLAIM 7.15

There is c_3 such that if D contains a $(c_0 + 2c_1)$-mountain, then $\chi(D) \leqslant c_3$.

Proof. Let Q be a $(c_0 + 2c_1)$-mountain of D. Then by Lemma 7.10, $|Q| \leqslant ((c_0 + 2c_1)!)^2$ and $\chi(Q) \geqslant c_0 + 2c_1$. For every partition Q into three sets Q_0, Q_1, Q_2, let Y_{Q_0,Q_1,Q_2} be the set of vertices $v \in V \backslash Q$ such that $Q_0 \subseteq N^o(v), Q_1 \subseteq N^+(v)$, and $Q_2 \subseteq N^-(v)$. Note that for every vertex $v \in V \backslash Q$, there always exists a partition of Q into some sets Q_0, Q_1, Q_2 such that v is non-adjacent with every vertex in Q_0, sees every vertex in Q_1 and is seen by every vertex in Q_2. Hence, $V \backslash Q$ can be written as the union of all possible Y_{Q_0,Q_1,Q_2}. There are $3^{|Q|}$ sets Y_{Q_0,Q_1,Q_2}.

(B) $\chi(Y_{Q_0,Q_1,Q_2}) \leqslant c_1$ for every partition (Q_0, Q_1, Q_2) of Q.

Indeed, if $Y_{Q_0,Q_1,Q_2} = \emptyset$, then (B) clearly holds. Otherwise, $Q_0 \subseteq N^o(v)$ for any $v \in Y_{Q_0,Q_1,Q_2}$. Note that the digraph Q_0 is $(H_1 \Rightarrow H_2)$-free and $(\alpha - 1)$-dense, and so by induction hypothesis, $\chi(Q_0) \leqslant c_0$. This gives $\chi(Q_1 \cup Q_2) \geqslant \chi(Q) - \chi(Q_0) \geqslant 2c_1$, implying that either $\chi(Q_1) \geqslant c_1$ or $\chi(Q_2) \geqslant c_1$. If $\chi(Q_1) \geqslant c_1$, then Q_1 contains a copy of H_2, say $\widehat{H}_2$. If $\chi(Y_{Q_0,Q_1,Q_2}) \geqslant c_1$, then Y_{Q_0,Q_1,Q_2} contains a copy of H_1, say $\widehat{H}_1$. Then $\widehat{H}_1 \Rightarrow \widehat{H}_2$ forms a copy of $H_1 \Rightarrow H_2$, a contradiction. Hence, $\chi(Y_{Q_0,Q_1,Q_2}) \leqslant c_1$. A similar argument establishes the case $\chi(Q_2) \geqslant c_1$. This proves (B).

Hence

$$\begin{aligned}\chi(D) &\leqslant \chi(Q) + \chi(V \setminus Q) \\ &\leqslant |Q| + \chi\Big(\bigcup_{(Q_0,Q_1,Q_2)} Y_{Q_0,Q_1,Q_2}\Big) \\ &\leqslant |Q| + \sum_{(Q_0,Q_1,Q_2)} \chi(Y_{Q_0,Q_1,Q_2}) \\ &\leqslant ((c_0+2c_1)!)^2 + 3^{((c_0+2c_1)!)^2} c_1.\end{aligned}$$

Set $c_3 := ((c_0+2c_1)!)^2 + 3^{((c_0+2b_1)!)^2} c_1$. This completes proof of the claim. □

Hence $\chi(D) \leqslant \max(c_2, c_3)$. Setting $f(\alpha) := \max(c_2, c_3)$ completes the proof of Lemma 7.14, thus proving Theorem 7.8. □

7.3 Cycles of superheroes

We will conclude this chapter with the proof of Theorem 7.9, which states that if H is a superhero, then so are $\Delta(H, T_k, T_1)$ and $\Delta(H, T_1, T_k)$ for any integer $k \geqslant 1$. We will prove that if H is a superhero, then so is $\Delta(H, T_k, T_1)$ for any $k \geqslant 1$. This is sufficient. Indeed, if H is a superhero, then so is H_{rev}, the digraph obtained from H by reversing all its arcs. Thus, $\Delta(H_{rev}, T_k, T_1)_{rev} = \Delta(H, T_1, T_k)$ is also a superhero.

THEOREM 7.16

For every superhero H and every pair of integers $k, \alpha \geqslant 1$, there is a number $f(H, k, \alpha)$ such that every α-dense $\Delta(H, T_k, T_1)$-free digraph D has $\chi(D) \leqslant f(H, k, \alpha)$.

Before starting the proof of Theorem 7.16, we would like to mention that T_k-free digraphs have bounded number of vertices, and so have bounded number of chromatic number. Indeed, it is proved in [Ste59] that for each integer $k \geqslant 1$, every tournament with at least 2^{k-1} vertices contains a copy of T_k. Let $\mathcal{R}(a, b)$ be the Ramsey number of (a, b), i.e., the smallest n such that any graph on n vertices either contains a stable set of order a or a clique of order b.

PROPOSITION 7.17

For each integer $k \geqslant 1$, every α-dense T_k-free digraph D has at most $\mathcal{R}(\alpha+1, 2^{k-1})$ vertices.

Proof. Suppose for a contradiction that there is a T_k-free digraph D with at least $\mathcal{R}(\alpha+1, 2^{k-1})$ vertices. Then the underlying graph of D contains either a stable set of size

$\alpha+1$ or a clique of size 2^{k-1}. The former case is impossible since D is α-dense. Thus D contains a tournament of size 2^{k-1}, and hence contains a copy of T_k, a contradiction. □

The idea of the proof of Theorem 7.16 is as follows. Fix a large number c and call a subset B of $V(D)$ with $\chi(B) = c$ a *bag*. We aim at finding a longest chain of disjoint bags $B_1, ..., B_t$ in $V(D)$, together with a partition $V(D)\setminus \bigcup B_i$ into sets we call *zones* $Z_0, ..., Z_t$ such that there is no *backward arc* in D (where uv is a backward arc if the bag or zone containing u has higher index than the bag or zone containing v). Then, using maximality of t, we show that the chromatic number of every zone is bounded. Once proving this, we observe that all B_i and Z_i have bounded chromatic number and since D has no backward arc, D has bounded chromatic number. However, the requirement that there is no backward arc in D is too strong and so we will have to slightly relax it. In doing so, we will need to allow some backward arcs, but we will want to do so in a very controlled manner. This leads us to the following definitions.

Given an integer $c \geqslant 1$ and an α-dense $\Delta(H, T_k, T_1)$-free digraph D, a set $B \subseteq V(D)$ with $\chi(B) = c$ is called a *c-bag*. A family of pairwise disjoint c-bags $B_1, ..., B_t$ is a *c-bag-chain* if for every i and every $v \in B_i$, we have $\chi(N^+(v) \cap B_{i-1}) \leqslant c_1$ and $\chi(N^-(v) \cap B_{i+1}) \leqslant c_1$, where c_1 is a fixed number satisfying

- $c_1 \geqslant \mathcal{R}(\alpha+1, 2^{k-1})$, and
- $c_1 \geqslant \chi(D)$ for every α-dense H-free digraph D.

Since H is a superhero, such c_1 clearly exists. Note that every α-dense T_k-free digraph D has at most c_1 vertices by Proposition 7.17, and so has chromatic number at most c_1.

Proof of Theorem 7.16. We proceed by induction on α. Since $\Delta(H, T_k, T_1)$ is a hero (see [BCC+13], Theorem 4.1), the theorem holds for $\alpha = 1$. Suppose that Theorem 7.16 is true for $\alpha - 1$ and let $c_0 = f(H, k, \alpha - 1)$. Let D be an α-dense $\Delta(H, T_k, T_1)$-free digraph with vertex set V. An important observation is that for every $v \in V$, the digraph $N^o(v)$ is $\Delta(H, T_k, T_1)$-free and is $(\alpha-1)$-dense, and hence

$$\chi(N^o(v)) \leqslant f(H, k, \alpha - 1) = c_0. \tag{7.1}$$

We also would like to recall some useful formulas. For every $v \in X \subseteq V$,

$$\chi(N^+(X)) \leqslant \sum_{v\in X} \chi(N^+(v)) \text{ and } \chi(N^o(X)) \leqslant \sum_{v\in X} \chi(N^o(v)), \tag{7.2}$$

and if $Y \subseteq V$ and $Y \cap X = \emptyset$, then (recalling that $M^+(X)$ is the set of vertices seen by all vertices of X)

$$Y = (M^+(X) \cap Y) \cup \big((N^-(X) \cup N^o(X)) \cap Y\big). \tag{7.3}$$

Let $|H| = h$ and set $c := 2(c_0 + c_1)(h + k)$. Let us assume that $B_1, ..., B_t$ is a c-bag-chain of D with t as large as possible. In the proof of this theorem, we drop prefix c- of c-bag and c-bag-chain for convenience. By definition of bag-chain, every bag has few backward arcs with bags preceeding or succeeding it. In the following claim, we show that every bag in fact has few backward arcs with any other bag.

CLAIM 7.18

For every i and $v \in B_i$, and for every $r > 0$,

(a) $\chi(N^+(v) \cap B_{i-r}) \leqslant c_1$, *and*

(b) $\chi(N^-(v) \cap B_{i+r}) \leqslant c_1$.

Proof. We proceed by induction on r. For $r = 1$, both (a) and (b) holds by definition of bag-chain. Suppose that both statements are true for $r - 1$. We now prove (a) for r. Suppose for a contradiction that there is v in some B_i such that $\chi(N^+(v) \cap B_{i-r}) \geqslant c_1$. Then $N^+(v) \cap B_{i-r}$ has a copy of H, say $\widehat{H}$. Then by applying (7.3) we have

$$B_{i-1} = \Big(M^+(\widehat{H}) \cap B_{i-1}\Big) \cup \Big(\big(N^-(\widehat{H}) \cup N^o(\widehat{H})\big) \cap B_{i-1}\Big),$$

and

$$B_{i-1} = \Big(N^-(v) \cap B_{i-1}\Big) \cup \Big(\big(N^+(v) \cup N^o(v)\big) \cap B_{i-1}\Big).$$

Thus (by using the fact that if $A = B \cup C = B' \cup C'$, then $A = (B \cap B') \cup C \cup C'$) we have

$$\begin{aligned} B_{i-1} = &\Big(M^+(\widehat{H}) \cap N^-(v) \cap B_{i-1}\Big) \cup \Big(\big(N^-(\widehat{H}) \cup N^o(\widehat{H})\big) \cap B_{i-1}\Big) \cup \\ &\cup \Big(\big(N^+(v) \cup N^o(v)\big) \cap B_{i-1}\Big). \end{aligned} \tag{7.4}$$

For each $x \in \widehat{H}$, by (7.1) we have $\chi(N^o(x) \cap B_{i-1}) \leqslant c_0$, and by induction hypothesis of (b) applied to x and $r - 1$, we have $\chi(N^-(x) \cap B_{i-1}) \leqslant c_1$. We also have $\chi(N^o(v) \cap B_{i-1}) \leqslant c_0$ by (7.1) and $\chi(N^+(v) \cap B_{i-1}) \leqslant c_1$ by definition of a bag-chain. Combining with (7.4) and (7.2) we have

$$\begin{aligned} \chi(M^+(\widehat{H}) \cap N^-(v) \cap B_{i-1}) &\geqslant \chi(B_{i-1}) - \sum_{x \in \widehat{H} \cup \{v\}} \chi(N^o(x) \cap B_{i-1}) \\ &\quad - \sum_{x \in \widehat{H}} \chi(N^-(x) \cap B_{i-1}) - \chi(N^+(v) \cap B_{i-1}), \\ &\geqslant 2(c_0 + c_1)(h + k) - c_0(h + 1) - c_1 h - c_1, \\ &\geqslant c_1. \end{aligned}$$

Then there exists a copy of T_k in $M^+(\widehat{H}) \cap N^-(v) \cap B_{i-1}$, say $\widehat{T}_k$. Note that by construction, we have all arcs from $\widehat{H}$ to $\widehat{T}_k$, from $\widehat{T}_k$ to v and from v to $\widehat{H}$. Then $\Delta(\widehat{H}, \widehat{T}_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction.

The proof of (b) for r is similar but not symmetric. In order to obtain a copy of $\Delta(H, T_k, T_1)$, we first get a copy of T_k in B_{i+r}, and then a copy of H in B_{i+1}. This proves the claim. □

We next prove a stronger statement that every bag has few backward arcs with the union of all other bags proceeding or succeeding it. Let $B_{i,j} = \bigcup_{s=i}^{j} B_s$ for any $1 \leqslant i \leqslant j \leqslant t$ (if $i < 1$ or $j > t$ or $j < i$, we set $B_{i,j} = \emptyset$).

CLAIM 7.19

For every i and $v \in B_i$,

- $\chi(N^+(v) \cap B_{1,i-2}) \leqslant c_1$, *and*
- $\chi(N^-(v) \cap B_{i+2,t})) \leqslant c_1$.

Proof. We repeat the same argument as in the proof of Claim 7.18. Suppose for a contradiction that the first statement is false, i.e., there is v in some B_i such that $\chi(N^+(v) \cap B_{1,i-2}) \geqslant c_1$. Then $N^+(v) \cap B_{1,i-2}$ has a copy of H, say $\widehat{H}$. For each $x \in \widehat{H}$, we have $\chi(N^o(x) \cap B_{i-1}) \leqslant c_0$ by (7.1), and $\chi(N^-(x) \cap B_{i-1}) \leqslant c_1$ by Claim 7.18. We also have $\chi(N^o(v) \cap B_{i-1}) \leqslant c_0$ and $\chi(N^+(v) \cap B_{i-1}) \leqslant c_1$. Thus by the same computation as in Claim 7.18, we obtain a copy of T_k in B_{i-1} and reach the contradiction. The proof of the second statement is similar. □

From Claim 7.19 we have the following immediate corollary.

CLAIM 7.20

For every i and $v \in B_i$,

- $\chi(N^+(v) \cap B_{1,i-1}) \leqslant 2c_1$.
- $\chi(N^-(v) \cap B_{i+1,t}) \leqslant 2c_1$.

We now show that the union of all bags has bounded chromatic number. We note that in the following proof, we will use only two hypotheses: Claim 7.20 and that $\chi(B_i)$ is bounded for every i. The reason for our remark is that we will re-use the arguments of this proof for subsequent claims.

CLAIM 7.21

$\chi(B_{1,t}) \leqslant 8cc_1$.

Proof. An arc uv with $u \in B_j, v \in B_i$, and $j > i$ is called a *backarc* with *span* $j - i$. For every i and every $v \in B_i$, if $|N^-(v) \cap B_{i+1,t}| < c_1 - 1$, let $F_v := N^-(u) \cap B_{i+1,t}$. If $|N^-(v) \cap B_{i+1,t}| \geqslant c_1 - 1$, let $F_v \subseteq N^-(v) \cap B_{i+1,t}$ consist of $c_1 - 1$ vertices whose backarcs to v have the $c_1 - 1$ largest spans. Let us show that

(A) For every backarc uv with $u \notin F_v$, if $u \in B_j$ and $v \in B_i$ then $\chi(B_{i,j}) \leqslant 4c$.

Indeed, if $j = i + 1$, then (A) clearly holds since $\chi(B_{i,i+1}) \leqslant \chi(B_i) + \chi(B_{i+1}) \leqslant 2c$. Thus we may suppose that $j \geqslant i + 2$. Since $u \in N^-(v) \cap B_{i+1,t}$ but $u \notin F_v$, it follows from the definition of F_v that $|F_v| = c_1 - 1$. Hence $|F_v \cup \{u\}| = c_1$. Then there is a copy of T_k in $F_v \cup \{u\}$, say $\widehat{T}_k$. Note that $\widehat{T}_k \subseteq B_{j,t}$.

We have a formula similar to (7.4):

$$B_{i+1,j-1} = \Big(M^-(\widehat{T}_k) \cap N^+(v) \cap B_{i+1,j-1}\Big) \cup \Big(\big(N^+(\widehat{T}_k) \cup N^o(\widehat{T}_k)\big) \cap B_{i+1,j-1}\Big) \cup \Big(\big(N^-(v) \cup N^o(v)\big) \cap B_{i-1}\Big).$$

For each $x \in \widehat{T}_k \cup \{v\}$, we have $\chi(N^o(x) \cap B_{i+1,j-1}) \leqslant c_0$. Note also that from Claim 7.20, we have $\chi(N^+(x) \cap B_{i+1,j-1}) \leqslant 2c_1$ for every $x \in \widehat{T}_k$ and $\chi(N^-(v) \cap B_{i+1,j-1}) \leqslant 2c_1$. Furthermore, if $\chi(M^-(\widehat{T}_k) \cap N^+(v) \cap B_{i+1,j-1}) \geqslant c_1$, then $M^-(\widehat{T}_k) \cap N^+(v) \cap B_{i+1,j-1}$ contains a copy of H, say $\widehat{H}$. Then $\Delta(\widehat{H}, \widehat{T}_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction. Hence

$$\begin{aligned}
\chi(B_{i+1,j-1}) &\leqslant \chi\Big(M^-(\widehat{T}_k) \cap N^+(v) \cap B_{i+1,j-1}\Big) + \sum_{x \in \widehat{T}_k \cup \{v\}} \chi\Big(N^o(x) \cap B_{i+1,j-1}\Big) \\
&\quad + \sum_{x \in \widehat{T}_k} \chi\Big(N^+(x) \cap B_{i+1,j-1}\Big) + \chi\Big(N^-(v) \cap B_{i+1,j-1}\Big) \\
&\leqslant c_1 + c_0(k+1) + 2c_1 k + 2c_1 \\
&\leqslant 3(c_1 + c_0)(k+1) \leqslant 2c,
\end{aligned}$$

since $c = 2(c_0 + c_1)(h + k)$. This gives $\chi(B_{i,j}) \leqslant \chi(B_i) + \chi(B_{i+1,j-1}) + \chi(B_j) \leqslant 4c$, which proves (A).

Now let G be the undirected graph with vertex set $B_{1,t}$ and $uv \in E(G)$ if $u \in F_v$ or $v \in F_u$. Then B_i is a stable set in G for every i. We now (properly) color the vertices of G by c_1 colors as follows. First, color all B_t properly by color 1. Suppose that we have already colored $B_{i+1}, ..., B_t$. Every vertex v in B_i is incident (in G) with at most $c_1 - 1$ vertices in $B_{i+1,t}$ (those belonging to F_v) and independent (in G) with all other vertices in B_i, so we can always properly color v, and so properly color B_i. When the process of coloring ends, we obtain a partition of $B_{1,t}$ into c_1 sets of colors, say $X_1, ..., X_{c_1}$, where each X_s is a stable set in G. We now claim that the chromatic number of the digraph induced by X_s is small.

(B) $\chi(X_s) \leqslant 8c$ for every $1 \leqslant s \leqslant c_1$.

To prove (B), we define a sequence of indices $i_1, i_2, \ldots$ inductively as follows. Let $i_1 = 1$, and for every $r \geqslant 1$, let $i_{r+1} > i_r$ be the smallest index such that $\chi(B_{i_r,i_{r+1}}) > 4c$. The sequence ends by i_ℓ with $\chi(B_{i_\ell,t}) \leqslant 4c$ (i.e. there is no $i_{\ell+1}$ satisfying the condition). Set $A_r := B_{i_r,i_{r+1}-1}$ for every $1 \leqslant r \leqslant \ell - 1$ and $A_\ell := B_{i_\ell,t}$. Then $B_{1,t} = \bigcup_{r=1}^{\ell} A_r$, and by definition of the sequence, $\chi(B_{i_r,i_{r+1}-1}) \leqslant 4c$ for every $1 \leqslant r \leqslant \ell - 1$. In other words, for every $1 \leqslant r \leqslant \ell$,

$$\chi(A_r) \leqslant 4c. \tag{7.5}$$

Suppose that there is a backarc uv with $u \notin F_v$ and $u \in A_r, v \in A_{r'}$, where $r \geqslant r' + 2$. Suppose that $u \in B_j$ and $v \in B_{j'}$. Then $j \geqslant i_r$ since $B_j \subseteq A_r = B_{i_r,i_{r+1}-1}$, and $j' < i_{r'+1}$ since $B_{j'} \subseteq A_{r'} = B_{i_{r'},i_{r'+1}-1}$, and so $j' < i_{r-1}$ since $r' + 1 \leqslant r - 1$. Also note that $\chi(B_{i_{r-1},i_r}) > 4c$ by definition of the sequence. Thus we have

$$\chi(B_{j',j}) \geqslant \chi(B_{i_{r-1},i_r}) > 4c,$$

which contradicts (A) that $\chi(B_{j',j}) \leqslant 4c$ if $u \notin F_v$. This implies the observation that for any $r \geqslant r' + 2$, there is no backarc uv with $u \in A_r, v \in A_{r'}$ and $u \notin F_v$.

Now fix an arbitrary X_s and let $X_{s,r} = X_s \cap A_r$ for every $1 \leqslant r \leqslant \ell$. Observe that if uv is an backarc with $u, v \in X_s$, then $u \notin F_v$ (otherwise, $u \in F_v$, so $uv \in G$ and so X_s is not stable in G, a contradiction). Hence combining with the observation in the paragraph above, we have that there is no backarc from $X_{s,r}$ to $X_{s,r'}$ for any r, r' with $r \geqslant r' + 2$; in other words, $X_{s,r'} \to X_{s,r}$ for any r, r' with $r \geqslant r' + 2$. By (7.5), we have that $\chi(X_{s,r}) \leqslant 4c$. Thus $\chi\big(\bigcup_{r\geqslant 1} X_{s,2r-1}\big) \leqslant 4c$ and $\chi\big(\bigcup_{r\geqslant 1} X_{s,2r}\big) \leqslant 4c$. This gives $\chi(X_s) = \chi\big(\bigcup_{r\geqslant 1} X_{s,r}\big) \leqslant 8c$, which proves (B).

Hence

$$\chi(B_{1,t}) \leqslant \sum_{s=1}^{c_1} \chi(X_s) \leqslant 8cc_1.$$

This completes the proof of Claim 7.21. □

We now turn our attention to the vertices not in the bags. We partition $V \setminus B_{1,t}$ into sets $Z_0, \ldots, Z_t$ called *zones* as follows. For every $x \in V \setminus B_{1,t}$, if i is the largest index such that $\chi(N^-(x) \cap B_i) \geqslant c_1$, then $x \in Z_i$; otherwise, $x \in Z_0$. We first show that every Z_i has few backward arcs with any B_j sufficiently far from it.

CLAIM 7.22

For every i and $v \in Z_i$,

- *$\chi(N^-(v) \cap B_{i+r}) \leqslant c_1$ for every $r \geqslant 1$, and*
- *$\chi(N^+(v) \cap B_{i-r}) \leqslant c_1$ for every $r \geqslant 2$.*

Proof. The former inequality is obvious by the partition criterion. For the latter, the proof follows the same idea as that of Claim 7.18, but is a bit more involved. Suppose for a contradiction that $\chi(N^+(v) \cap B_{i-r}) \geqslant c_1$ for some $r \geqslant 2$. Then there is a copy of H in $N^+(v) \cap B_{i-r}$, say $\widehat{H}$. Since $\chi(N^-(v) \cap B_i) \geqslant c_1$ by partition criterion, there is a copy of T_k in $N^-(v) \cap B_i$, say $\widehat{T}_k$. Then by applying (7.3) we have

$$B_{i-1} = \Big(M^+(\widehat{H}) \cap B_{i-1}\Big) \cup \Big(\big(N^-(\widehat{H}) \cup N^o(\widehat{H})\big) \cap B_{i-1}\Big),$$

$$B_{i-1} = \Big(M^-(\widehat{T}_k) \cap B_{i-1}\Big) \cup \Big(\big(N^+(\widehat{T}_k) \cup N^o(\widehat{T}_k)\big) \cap B_{i-1}\Big),$$

and

$$B_{i-1} = \Big(\big(N^+(v) \cup N^-(v)\big) \cap B_{i-1}\Big) \cup \Big(N^o(v) \cap B_{i-1}\Big).$$

Let $R = M^+(\widehat{H}) \cap M^-(\widehat{T}_k) \cap \big(N^+(v) \cup N^-(v)\big) \cap B_{i-1}$. Then we have

$$\begin{aligned} B_{i-1} = R \cup \Big(\big(\, N^-(\widehat{H}) \cup N^o(\widehat{H})\big) \cap B_{i-1}\Big) \cup \\ \cup \Big(\big(N^+(\widehat{T}_k) \cup N^o(\widehat{T}_k)\big) \cap B_{i-1}\Big) \cup \big(N^o(v) \cap B_{i-1}\big). \end{aligned}$$

For each $x \in \widehat{H} \cup \widehat{T}_k \cup \{v\}$, we have $\chi(N^o(x) \cap B_{i-1}) \leqslant c_0$. By Claim 7.18, we have $\chi(N^-(x) \cap B_{i-1}) \leqslant c_1$ for each $x \in \widehat{H}$ and $\chi(N^+(x) \cap B_{i-1}) \leqslant c_1$ for each $x \in \widehat{T}_k$. Then

$$\begin{aligned} \chi(R) &\geqslant \chi(B_{i-1}) - \sum_{x \in \widehat{H} \cup \widehat{T}_k \cup \{v\}} \chi(N^o(x) \cap B_{i-1}) \\ &\quad - \sum_{x \in \widehat{H}} \chi(N^-(x) \cap B_{i-1}) - \sum_{x \in \widehat{T}_k} \chi(N^+(x) \cap B_{i-1}) \\ &\geqslant 2(c_0 + c_1)(h + k) - c_0(h + k + 1) - c_1 h - c_1 k \\ &\geqslant 2c_1. \end{aligned}$$

Let $R_1 = R \cap N^+(v)$ and $R_2 = R \cap N^-(v)$. Note that $R = R_1 \cup R_2$, and so either $\chi(R_1) \geqslant c_1$ or $\chi(R_2) \geqslant c_1$. If $\chi(R_1) \geqslant c_1$, there is a copy of H in R_1, say $\widehat{H}'$. Then $\Delta(\widehat{H}', \widehat{T}_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction. If $\chi(R_2) \geqslant c_1$, there is a copy of T_k in R_2, say $\widehat{T}'_k$. Then $\Delta(\widehat{H}, \widehat{T}'_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction again. This proves the claim. □

While Claim 7.22 shows that every zone has few backward arcs with every sufficiently far bag. The next claim shows the converse, i.e., every bag has few backward arcs with every sufficiently far zone.

CLAIM 7.23

For every i and $v \in B_i$,

- *$\chi(N^+(v) \cap Z_{i-r}) \leqslant c_1$ for every $r \geqslant 2$, and*
- *$\chi(N^-(v) \cap Z_{i+r}) \leqslant c_1$ for every $r \geqslant 3$.*

Proof. We repeat the argument in the proof of Claim 7.18. Suppose for a contradiction that the first statement is false, i.e., there is v in some B_i and $r \geqslant 2$ such that $\chi(N^+(v) \cap Z_{i-r}) \geqslant c_1$. Then $N^+(v) \cap Z_{i-r}$ has a copy of H, say $\widehat{H}$. For each $x \in \widehat{H}$, we have $\chi(N^o(x) \cap B_{i-1}) \leqslant c_0$ by (7.1), and $\chi(N^-(x) \cap B_{i-1}) \leqslant c_1$ by the first inequality in Claim 7.22. We also have $\chi(N^o(v) \cap B_{i-1}) \leqslant c_0$ and $\chi(N^+(v) \cap B_{i-1}) \leqslant c_1$ by definition of bag-chain. Thus by the same computation as in Claim 7.18, we obtain a copy of T_k in B_{i-1} and reach the contradiction. The proof of the second statement is similar. □

The next claim is a counterpart of Claim 7.20 for zones, and will be used to bound the chromatic number of the union of zones.

CLAIM 7.24

For every i and $v \in Z_i$,

- $\chi(N^+(v) \cap \bigcup_{s=0}^{i-3} Z_s) \leqslant c_1$, *and*
- $\chi(N^-(v) \cap \bigcup_{s=i+3}^{t} Z_s) \leqslant c_1$.

Proof. Suppose for a contradiction that $\chi(N^+(v) \cap \bigcup_{s=0}^{i-3} Z_s) \geqslant c_1$, then there is a copy of H in $N^+(v) \cap \bigcup_{s=0}^{i-3} Z_s$, say $\widehat{H}$. For each $x \in \widehat{H}$, we have $\chi(N^o(x) \cap B_{i-2}) \leqslant c_0$ by (7.1), and $\chi(N^-(x) \cap B_{i-2}) \leqslant c_1$ by the first inequality in Claim 7.22. We also have $\chi(N^o(v) \cap B_{i-2}) \leqslant c_0$ and $\chi(N^+(v) \cap B_{i-2}) \leqslant c_1$ by the second inequality in Claim 7.22. Thus by the same computation as in Claim 7.18, we obtain that $\chi(M^+(\widehat{H}) \cap N^-(v) \cap B_{i-2}) \geqslant c_1$ implying that there is a copy of T_k in $M^+(\widehat{H}) \cap N^-(v) \cap B_{i-2}$, reaching a contradiction. The proof of the second statement is similar, where we use B_{i+1} instead of B_{i-2}. □

In order to bound the chromatic number of the union of zones, we also need that each zone has bounded chromatic number. We will prove this by employing the assumption that the bag-chain $B_1, ..., B_t$ is of maximum length.

CLAIM 7.25

No zone Z_i contains a bag-chain of length 6.

Proof. Suppose that some Z_i contains a bag-chain of length 6, say $Y_1, ..., Y_6$. Note that we have

- $\chi(N^+(v) \cap B_{i-3}) \leqslant c_1$ for every $v \in Y_1$ by Claim 7.22 and $\chi(N^-(v) \cap Y_1) \leqslant c_1$ for every $v \in B_{i-3}$ by Claim 7.23; and

- $\chi(N^-(v) \cap B_{i+3}) \leqslant c_1$ for every $v \in Y_6$ by Claim 7.22 and $\chi(N^+(v) \cap Y_6) \leqslant c_1$ for every $v \in B_{i+3}$ by Claim 7.23.

Thus by definition of bag-chain, $B_1, ..., B_{i-3}, Y_1, Y_2, ..., Y_6, B_{i+3}, ..., B_t$ is a bag-chain of length $t+1$, which contradicts the maximality of t. Hence no Z_i contains a bag-chain of length 6. □

LEMMA 7.26

There is c' such that if an α-dense $\Delta(H, T_k, T_1)$-free digraph D' does not contain any c-bag-chain of length 6, then $\chi(D') \leqslant c'$.

We defer the proof of Lemma 7.26 for now.

We now show that this is sufficient to prove the theorem. To do so, we group zones by indices modulo 3 and follow a similar argument as in Claim 7.21. Fixed $0 \leqslant s \leqslant 2$, and for every $i \equiv s \mod 3$, $1 \leqslant i \leqslant t$, let $Z^s_{\lfloor i/3 \rfloor} := Z_i$. Then $\chi(Z^s_j) \leqslant c'$ for every j, and for every $v \in Z^s_j$, by Claim 7.24, we have

- $\chi(N^+(v) \cap \bigcup_{r<j} Z^s_r) \leqslant c_1$, and
- $\chi(N^-(v) \cap \bigcup_{r>j} Z^s_r) \leqslant c_1$.

By applying Lemma 7.26 for $D' = Z_i$ and using Claim 7.25, it is immediate that $\chi(Z_i) \leqslant c'$ for every i. Thus, $\chi(Z^s_j) \leqslant c'$ for every j. We can repeat exactly the argument of Claim 7.21 (with c replaced by c' and by using Claim 7.24 instead of Claim 7.20) to deduce

$$\chi\Big(\bigcup_{r\geqslant 0} Z^s_r\Big) \leqslant 8c'c_1$$

for every $s = 0, 1, 2$. (Remark: we may suppose that $c' \geqslant c$, which is necessary for the argument at the end of the proof of Property (A) that $3(c_1 + c_0)(k+1) \leqslant 2c \leqslant 2c'$).

Hence

$$\chi(\bigcup_{s=0}^{t} Z_s) \leqslant \sum_{s=0}^{2} \chi\Big(\bigcup_{r\geqslant 0} Z^s_r\Big) \leqslant 24c'c_1.$$

From Claims 7.21 and the above inequality, we have

$$\chi(D) \leqslant \chi(B_{1,t}) + \chi(\cup_{i=0}^{t} Z_i) \leqslant 8cc_1 + 24c'c_1,$$

which proves Theorem 7.16. □

Proof of Lemma 7.26

Lemma 7.26 asserts that the absence of bag-chain of length 6 is enough to force a digraph (α-dense and $\Delta(H, T_k, T_1)$-free) to have bounded chromatic number. It is trivial by definition that the absence of bag-chain of length 1 forces bounded chromatic number. In the next lemma, we show that the absence of bag-chain of length 2 forces bounded chromatic number, which contains most of the difficulty of the proof of Lemma 7.26. In the following lemma, c, c_0, c_1, h, k are the values as in the proof of Theorem 7.16. Recall that $c = 2(c_0 + c_1)(h + k)$.

LEMMA 7.27

For every $d \geqslant c$, there is $g(d)$ such that if an α-dense $\Delta(H, T_k, T_1)$-free digraph D contains no d-bag-chain of length 2, then $\chi(D) \leqslant g(d)$.

Proof. Let J be the tournament $H \Rightarrow T_k$. By Theorem 7.8, J is a superhero and contains both H and T_k as subtournaments. Let D be a digraph with vertex set V and satisfying the hypotheses of the lemma. A copy of J in D is called a *ball*. A ball $\widehat{J}$ is colored *red* if $\chi(M^+(\widehat{J})) \leqslant d$ and *blue* if $\chi(M^-(\widehat{J})) \leqslant d$ (note that we do not color vertices of a ball but color the ball as a single object). A ball certainly can be both red and blue, in which case we color it arbitrarily with one of the colors.

CLAIM 7.28

Every ball is either red or blue.

Proof. Suppose for a contradiction that a ball $\widehat{J}$ is neither red nor blue. Then there are $B_1 \subseteq M^-(\widehat{J})$ and $B_2 \subseteq M^+(\widehat{J})$ such that $\chi(B_1) = d$ and $\chi(B_2) = d$. Suppose that $\chi(N^+(v) \cap B_1) \geqslant c_1$ for some $v \in B_2$. Let $\widehat{H}$ be a copy of H in $N^+(v) \cap B_1$. Let $\widehat{T}_k$ be a copy of T_k in $\widehat{J}$. Then $\Delta(\widehat{H}, \widehat{T}_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction. Hence $\chi(N^+(v) \cap B_1) \leqslant c_1$ for every $v \in B_2$. Similarly, $\chi(N^-(v) \cap B_2) \leqslant c_1$ for every $v \in B_1$, for otherwise a copy of T_k in $N^-(v) \cap B_2$ would yield a contradiction. Hence B_1, B_2 is a d-bag-chain of length 2, a contradiction. □

For every vertex v of V, we color v as follows. If there are $c_1 + 1$ vertex-disjoint red balls $\widehat{J}_1, ..., \widehat{J}_{c_1+1}$ such that $\widehat{J}_i$ has complete arcs to $\widehat{J}_j$ for every $i < j$, and $v \in \widehat{J}_{c_1+1}$, then we color v *red*. If there are $c_1 + 1$ blue balls $\widehat{J}_1, ..., \widehat{J}_{c_1+1}$ such that $\widehat{J}_i$ has complete arcs to $\widehat{J}_j$ for every $i < j$, and $v \in \widehat{J}_1$, then we color v *blue*. If v satisfies both conditions, we color v arbitrarily. After the process of coloring, we obtain a partition of V into R the set of red vertices, B the set of blue vertices, and U the set of uncolored vertices.

CLAIM 7.29

There is d_1 such that $\chi(U) \leqslant d_1$.

Proof. Let K be the tournament of $J \Rightarrow J \Rightarrow ... \Rightarrow J$ ($2c_1 + 2$ times J). Since J is a superhero, then so is K by Theorem 7.8. Hence there is d_1 such that every α-dense K-free digraph D' has chromatic number at most d_1. Suppose that U contains a copy of K, say $\widehat{K}$. Since every ball is either red or blue, we can find $c_1 + 1$ vertex-disjoint monochromatic balls $\widehat{J}_1, ..., \widehat{J}_{c_1}$ in $\widehat{K}$ such that $\widehat{J}_i$ has complete arcs to $\widehat{J}_j$ for every $i < j$. Then either vertices of $\widehat{J}_1$ are blue or vertices of $\widehat{J}_{c_1+1}$ are red, a contradiction with the fact that all vertices of U are uncolored. Hence U is K-free, and so $\chi(U) \leqslant d_1$. □

It remains to show that R and B have bounded chromatic number, which can be done by applying Theorems 6.14. To do so, we need to prove that $N^+(v)$ has bounded chromatic number for every $v \in R$.

CLAIM 7.30

There is d_2 such that $\chi(N^+(v)) \leqslant d_2$ for every $v \in R$.

Proof. Fix $v \in R$. Then there are vertex-disjoint red balls $\widehat{J}_1, ..., \widehat{J}_{c_1+1}$ where $v \in \widehat{J}_{c_1+1}$ and $\widehat{J}_i$ has complete arcs to $\widehat{J}_j$ for every $i < j$. Let $L = \bigcup_{i=1}^{c_1} \widehat{J}_1$. Note that v is seen by all vertices of L. For every $u \in L$, we have $\chi(N^+(v) \cap N^o(u)) \leqslant c_0$. Hence

$$\chi(N^+(v) \cap N^o(L)) \leqslant \sum_{u \in L} \chi(N^+(v) \cap N^o(u)) \leqslant |L| c_0.$$

For every partition of L into L_1, L_2 (L_1 or L_2 may be empty), let $Y_{L_1,L_2} = N^+(v) \cap M^-(L_1) \cap M^+(L_2)$. Observe that for every vertex $x \in N^+(v) \backslash N^o(L)$ (note that $x \notin L$ since $v \in \widehat{J}_{c_1+1}$), there is a partition of L into some L_1, L_2 such that x sees all vertices of L_1 and is seen by all vertices of L_2, and so $x \in Y_{L_1,L_2}$. Hence

$$N^+(v) \backslash N^o(L) = \bigcup_{(L_1,L_2)} Y_{L_1,L_2}.$$

We now show that $\chi(Y_{L_1,L_2}) \leqslant d$ for every Y_{L_1,L_2}.

- If, for some $1 \leqslant i \leqslant c_1$, there is $\widehat{J}_i \subseteq L_2$, recall that $\chi(M^+(\widehat{J}_i)) \leqslant d$ since $\widehat{J}_i$ is red. Since $Y_{L_1,L_2} \subseteq M^+(L_2) \subseteq M^+(\widehat{J}_i)$, we have that $\chi(Y_{L_1,L_2}) \leqslant d$.

- Otherwise, if there is no $\widehat{J}_i \subseteq L_2$, then L_1 contains at least one vertex of each $\widehat{J}_i$, $1 \leqslant i \leqslant c_1$, and so $|L_1| \geqslant c_1$. Hence L_1 has a copy of T_k, say $\widehat{T}_k$. Note that all vertices of $\widehat{T}_k$ see v since $v \in \widehat{J}_{c_1+1}$. If $\chi(Y_{L_1,L_2}) \geqslant c_1$, then Y_{L_1,L_2} contains a copy of H, say $\widehat{H}$, then $\Delta(\widehat{H}, \widehat{T}_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction. Hence $\chi(Y_{L_1,L_2}) \leqslant c_1 \leqslant d$.

Note that $|L| = |J|c_1 = (h+k)c_1$. Thus, there are $2^{(h+k)c_1}$ possible ways to partition L into L_1, L_2. Hence

$$\begin{aligned}\chi(N^+(v)) &\leqslant \chi\Big(N^+(v) \cap N^o(L)\Big) + \chi\Big(N^+(v) \backslash N^o(L)\Big) \\ &\leqslant |L|c_0 + \sum_{(L_1,L_2)} \chi(Y_{L_1,L_2}) \\ &\leqslant (h+k)c_1c_0 + 2^{(h+k)c_1}d.\end{aligned}$$

Set $d_2 := (h+k)c_1c_0 + 2^{(h+k)c_1}d$. This completes the proof of the claim. □

Then $\chi(N^+(v) \cap R) \leqslant \chi(N^+(v)) \leqslant d_2$ for every $v \in R$. Then by applying Theorem 6.14 for digraph R with $t = d_2$, we have $\chi(R) \leqslant d_3$ for some d_3.

We now prove that B has bounded chromatic number. The proof is slightly different from that of Claim 7.30 due to asymmetry of $\Delta(H, T_k, T_1)$.

CLAIM 7.31

There is d_4 such that $\chi(N^-(v)) \leqslant d_4$ for every $v \in B$.

Proof. Fix $v \in B$. Then there are vertex-disjoint blues balls $\widehat{J}_1, ..., \widehat{J}_{c_1+1}$ where $v \in \widehat{J}_1$ and $\widehat{J}_i$ has complete arcs to $\widehat{J}_j$ for every $i < j$. Let $L = \bigcup_{i=2}^{c_1+1} \widehat{J}_i$. Note that v sees all vertices of L. For every $u \in L$, we have

$$\chi(N^-(v) \cap N^o(L)) \leqslant \sum_{u \in L} \chi(N^-(v) \cap N^o(u)) \leqslant |L|c_0.$$

For every partition of L into L_1, L_2 (L_1 or L_2 may be empty), let $Y_{L_1,L_2} = N^-(v) \cap M^-(L_1) \cap M^+(L_2)$. Observe that for every vertex $x \in N^-(v) \backslash N^o(L)$ (note that $x \notin L$ since $v \in \widehat{J}_1$), there is a partition of L into some L_1, L_2 such that x sees all vertices of L_1 and is seen by all vertices of L_2, and so $x \in Y_{L_1,L_2}$. Hence

$$N^-(v) \backslash N^o(L) = \bigcup_{(L_1,L_2)} Y_{L_1,L_2}.$$

We now show that $\chi(Y_{L_1,L_2}) \leqslant d$ for every Y_{L_1,L_2}.

- If there is $\widehat{J}_i \subseteq L_2$, then $\widehat{J}_i$ contains a copy of H, say $\widehat{H}$. Note that v is in the first ball, so v sees all vertices of $\widehat{H}$. If $\chi(Y_{L_1,L_2}) \geqslant c_1$, then Y_{L_1,L_2} contains a copy of T_k, say $\widehat{T}_k$, then $\Delta(\widehat{H}, \widehat{T}_k, v)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction. Hence $\chi(Y_{L_1,L_2}) \leqslant c_1 \leqslant d$.
- If $\widehat{J}_{c_1+1} \subseteq L_1$, recall that $\chi(M^-(\widehat{J}_{c_1+1})) \leqslant d$ since $\widehat{J}_{c_1+1}$ is blue. Since $Y_{L_1,L_2} \subseteq M^-(L_1) \subseteq M^-(\widehat{J}_{c_1+1})$, we have that $\chi(Y_{L_1,L_2}) \leqslant d$.

- Otherwise, we have two remarks:

  (i) $\widehat{J}_{c_1+1}$ must have a vertex in L_2, say z, and

  (ii) (2) Every $\widehat{J}_i, 2 \leqslant i \leqslant c_1$ must have a vertex in L_1.

  Then $|L_1 \cup \{v\}| \geqslant c_1$, and so $L_1 \cup \{v\}$ contains a copy of T_k, say $\widehat{T}'_k$, such that all vertices of $\widehat{T}'_k$ are in one of the $\widehat{J}_i$, $1 \leqslant i \leqslant c_1$ (note that $v \in \widehat{J}_1$). Observe that z is seen by all vertices of $\widehat{T}'_k$ since $z \in \widehat{J}_{c_1+1}$. If $\chi(Y_{L_1,L_2}) \geqslant c_1$, then Y_{L_1,L_2} contains a copy of H, say $\widehat{H}'$, then $\Delta(\widehat{H}', \widehat{T}'_k, z)$ forms a copy of $\Delta(H, T_k, T_1)$, a contradiction. Hence $\chi(Y_{L_1,L_2}) \leqslant c_1 \leqslant d$.

Hence using a computation similar to the claim above, we have $\chi(N^-(v)) \leqslant d_4$, where $d_4 := (h+k)c_1c_0 + 2^{(h+k)c_1}d$. This completes the proof of the claim. □

Then $\chi(N^-(v) \cap B) \leqslant \chi(N^-(v)) \leqslant d_4$ for every $v \in B$. Then by applying Theorem 6.18 to the digraph B with $t = d_4$, we have $\chi(B) \leqslant d_5$ for some d_5. Hence

$$\chi(D) \leqslant \chi(B) + \chi(R) + \chi(U) \leqslant d_1 + d_3 + d_5.$$

This completes the proof of Lemma 7.27. □

We are now ready to prove Lemma 7.26.

CLAIM 7.32

> *If an α-dense $\Delta(H, T_k, T_1)$-free digraph D contains no c-bag-chain of length 6, then $\chi(D) \leqslant g(g(g(c)))$ where g is the function in Lemma 7.27.*

Proof. Suppose for a contradiction that $\chi(Z_i) \geqslant g(g(g(c)))$. We will show that D contains a c-bag-chain of length 8. By applying Lemma 7.27 to D with $d := g(g(c))$, we have that D contains a $g(g(c))$-bag-chain of length two, say X_1, X_2. Hence $\chi(X_1) = \chi(X_2) = g(g(c))$ and

- $\chi(N^+(v) \cap X_1) \leqslant c_1$ for every $v \in X_2$, and
- $\chi(N^-(v) \cap X_2) \leqslant c_1$ for every $v \in X_1$.

Apply Lemma 7.27 again to X_1 (res. X_2) with $d := g(c)$, we obtain a $g(c)$-bag-chain of length two, say Y_1, Y_2 in X_1 (res. Y_3, Y_4 in X_2). Since $Y_2 \subset X_1$ and $Y_3 \subset X_2$, we have

- $\chi(N^+(v) \cap Y_2) \leqslant c_1$ for every $v \in Y_3$, and
- $\chi(N^-(v) \cap Y_3) \leqslant c_1$ for every $v \in Y_2$.

Hence by definition, $Y_1, ..., Y_4$ forms a $g(c)$-bag-chain of length 4. Note that $\chi(Y_s) = g(c)$ for every $1 \leqslant s \leqslant 4$. Repeating the argument we obtain a c-bag-chain of length 2 inside each Y_s, and hence obtain a c-bag-chain $B_1, ..., B_8$ of length 8 inside D. This contradicts the fact that D has no c-bag-chain of length 6, and so completes the proof. □

Claim 7.32 proves Lemma 7.26, concluding the proof of Theorem 7.16.

Chapter 8

Dense Triangle-Free Digraphs

8.1 Coloring triangle-free digraphs

This chapter is devoted to investigate a special case, where the pattern is the directed triangle C_3. The aim of this chapter is to present a potential direction to prove that C_3 has the general Erdős-Hajnal property (recall that no nontrivial pattern, even C_3, is known to have the general property). We believe that the following conjecture, which is a stronger statement than the general property of C_3, is true.

CONJECTURE 8.1

> *There is an integer ℓ such that for every positive integer α, if D is an α-dense C_3-free digraph, then D can be properly colored by α^ℓ colors.*

Indeed, if Conjecture 8.1 is true, then every C_3-free digraph D either has a stable (hence, acyclic) set of size $n^{1/2\ell}$ or can be colored by $(n^{1/2\ell})^\ell = \sqrt{n}$ colors, and hence has an acyclic set of size $n/\sqrt{n} = \sqrt{n}$. Thus C_3 has the general Erdős-Hajnal property.

While targeting a polynomial bound for the number of colors in Conjecture 8.1, we could not even achieve an exponential bound. However, by an algorithmic approach, we are able to obtain a factorial bound, and the coloring process can be done in polynomial time.

The content of this chapter is covered in parts of papers *Coloring dense digraphs*, Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé, *Combinatorica*, to appear, and *Domination and fractional domination in digraphs*, Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé, *Electronic Journal of Combinatorics*, to appear.

THEOREM 8.2

For every positive integer α, if D is an α-dense C_3-free digraph, then D can be properly colored by $35^{\alpha-1}\alpha!$ colors, and such a coloring can be found in polynomial time.

In this section, we prove Theorem 8.2. We present an efficient algorithm to color every α-dense C_3-free digraph with at most $35^{\alpha-1}\alpha!$ colors. For a digraph D, let $n = |V(D)|$ denote the size of its vertex set. Let $poly(n)$ denote the function n^k for some rational number $k > 0$.

For each integer $\alpha \geqslant 1$, define $h(\alpha)$ to be the minimum number such that every α-dense C_3-free digraph D has chromatic number at most $h(\alpha)$. Conjecture 8.1 asserts that $h(\alpha) \leqslant \alpha^{\ell}$ for some ℓ. Clearly $h(1) = 1$ since every C_3-free tournament is acyclic. However, $h(\alpha)$ is still unknown for all $\alpha \geqslant 2$. We believe that $h(2) = 2$ or 3 even though the best bound we have is $h(2) \leqslant 18$ (by tweaking the proof of Theorem 8.2). Theorem 8.2 gives an factorial upper bound for $h(\alpha)$ by the function $g(\alpha) := 35^{\alpha-1}\alpha!$. Since our proof of Theorem 8.2 will use induction, we will assume that every $(\alpha-1)$-dense C_3-free digraph can be colored with at most $35^{\alpha-2}(\alpha-1)!$ colors. This is true for $\alpha = 1$, and we will prove it for $\alpha > 1$ by induction.

We begin with some observations regarding the size of a dominating set in a digraph.

PROPOSITION 8.3

A digraph D has an acyclic dominating set, and this set can be found in time $poly(n)$.

Proof. We proceed by induction on n. The statement clearly holds for $n = 1$. For $n > 1$, pick an arbitrary vertex v. Then $V(D) \setminus v = N^o(v) \cup N^-(v) \cup N^+(v)$. Applying induction to the subgraph, $D[N^o(v) \cup N^-(v)]$, we obtain an acyclic dominating set S'. Then $S := S' \cup \{v\}$ is a dominating set of D. Note that $S' \subseteq N^o(v) \cup N^-(v)$, so v does not see any vertex of S'. Hence S is an acyclic set since S' is an acyclic set. The running time for this procedure is $poly(n)$. □

PROPOSITION 8.4

Given an α-dense C_3-free digraph D, there is a set $Y \subseteq V(D)$ of size at most α such that $V(D) = Y \cup N^o(Y) \cup N^+(Y)$, and this set can be found in time $poly(n)$.

Proof. First apply Proposition 8.3 to obtain an acyclic dominating set S of D. Then apply Proposition 6.16 to $D[S]$ to obtain a stable dominating set Y of $D[S]$ of size at most $\alpha(D)$ in time $poly(n)$.

We now show that Y is a set with the desired properties. Suppose for a contradiction that there is $v \notin Y \cup N^o(Y) \cup N^+(Y)$. Then $Y \subseteq N^+(v)$ and $v \notin S$ since Y dominates all vertices of S. There is $u \in S$ seeing v since S is a dominating set of D. Note that $u \notin Y$; otherwise this contradicts $Y \subseteq N^+(v)$. There is $y \in Y$ seeing u since Y is a dominating set of S. Then u, v, y are distinct vertices where u sees v, v sees y, and y sees u. Hence we obtain a copy of C_3 in D, a contradiction. □

We now present some definitions and useful lemmas. First, we re-define a *bag* so that it inherits all properties of a bag as defined in Section 7.3 and so that it can be tested efficiently.

DEFINITION 8.5

For a digraph D, we say that $B \subseteq V(D)$ is a bag of D, if every three distinct vertices $\{x, y, z\} \in V(D) \setminus B$ have a common neighbor in B.

Recall that u and v are neighbors if either uv or vu is an arc. We can check in $poly(n)$ time (e.g. $O(n^4)$) whether or not a set B is a bag of D by exhaustively checking all triples in $V(D) \setminus B$. Suppose that the n vertices of a digraph can be partitioned into disjoint sets such that the subgraph induced on each set is $(\alpha - 1)$-dense. Let $time(\alpha - 1, n)$ denote the maximum (over all such possible partitions of the vertices) total time required by our algorithm (the algorithm COLOR-DIGRAPH, which we will define shortly) to color all of the subgraphs, each with at most $35^{\alpha-2}(\alpha - 1)!$ colors by induction. The following claim follows from this definition.

CLAIM 8.6

Suppose $\sum_{i=1}^{\ell} n_i = n$, where $n_i \geqslant 1$ is an integer. Then $\sum_{i=1}^{\ell} time(\alpha, n_i) \leqslant time(\alpha, n)$.

We now fix an arbitrary α-dense C_3-free digraph D, and we omit the subscript D from the relevant notation when the context is clear. To simplify, let us set $\beta = 35^{\alpha-2}(\alpha - 1)!$.

CLAIM 8.7

If $S \subset V(D)$ is not a bag of D, then:

(a) We can partition S into three disjoint sets, S_1, S_2 and S_3, each is $(\alpha - 1)$-dense. This procedure takes time $poly(n)$.

(b) We can color S with 3β colors in time $time(\alpha - 1, |S|)$.

Proof. If S is not a bag of D, then we can, in time $poly(n)$, find a triple $\{x, y, z\} \in V(D) \setminus S$ such that every $v \in S$ is not incident to at least one of x, y or z. Thus, each vertex $v \in S$ belongs to either $N^o(x) \cap S, N^o(y) \cap S$ or $N^o(z) \cap S$. Each of these sets is $(\alpha - 1)$-dense. The total time for this procedure is $poly(n)$. The second assertion follows from the (a) and from the definition of the function $time(\alpha, n)$. □

DEFINITION 8.8

A bag $B \subseteq V(D)$ is poor *if for every vertex $v \in B$, either $N^-(v) \cap B$ or $N^+(v) \cap B$ is not a bag of D.*

We can check in $poly(n)$ time (e.g. $O(n^5)$) if a bag B is poor by testing whether or not $N^-(v) \cap B$ and $N^+(v) \cap B$ are bags for every $v \in B$.

CLAIM 8.9

If $B \subseteq V(D)$ is a poor bag, then we can color B with $8\alpha\beta$ colors in time $poly(n) + time(\alpha - 1, |B|)$.

Proof. Since B is poor, then for each $v \in B$, either $N^-(v) \cap B$ or $N^+(v) \cap B$ is not a bag of D. Then we can partition B into two sets, L and R, where $N^-(v) \cap B$ is not a bag for every $v \in L$, and $N^+(v) \cap B$ is not a bag for every $v \in R$.

Applying Proposition 8.4 to R, we can find a set $Y_R \subseteq R$ such that $|Y_R| \leqslant \alpha$ and $R \subseteq Y_R \cup N^+(Y_R) \cup N^o(Y_R)$. And so:

$$R = \bigcup_{v \in Y_R} \Big(\big(N^+(v) \cup N^o(v) \cup \{v\}\big) \cap R \Big).$$

For $v \in Y_R$, note that $N^+(v) \cap R$ is not a bag, and so by (a) from Claim 8.7, we can partition $N^+(v) \cap R$ into three $(\alpha - 1)$-dense sets. Additionally, we have the set $N^o(v) \cap R$ is also $(\alpha - 1)$-dense. Overall, we can partition $R \setminus Y_R$ into 4α sets, each is $(\alpha - 1)$-dense.

Similarly, one can find a set $Y_L \subseteq L$ such that $|Y_L| \leqslant \alpha$ and $L \setminus Y_L$ can be partitioned into $4\alpha(\alpha - 1)$-dense sets. Therefore, in time $poly(n)$, we can partition $B \setminus Y_R \cup Y_L$, and in time $time(\alpha - 1, |B|)$ we can color $B \setminus \{Y_r \cup Y_L\}$ with $8\alpha\beta$ colors. We can then color each $v \in Y_R$ (respectively, $v \in Y_L$) with an arbitrary color used to color the set $N^o(v) \cap R$ (respectively, $N^o(v) \cap L$). □

In general, we do not know how to color a bag efficiently, and a bag may be very large (e.g. $V(D)$ is a bag). Our aim is therefore to find poor bags, since these can be colored using Claim 8.9. The first step of our algorithm is to find a chain of poor bags.

DEFINITION 8.10

> *A sequence of pairwise disjoint bags $B_1, \ldots, B_t$ forms a* chain of bags *if $B_i \to B_{i+1}$ for every $i \in [1, t)$.*

Recall that $B_{i+1} \to B_i$ means there is no arc from B_i to B_{i+1}. Moreover, if each B_i is a poor bag, then this sequence is a *chain of poor bags*. Given a chain of bags $C = \{B_1, B_2, \ldots, B_t\}$ for D, we say that $v \in C$ if $v \in B_i$ for some $i \in [1, t]$. We can partition the vertices in $V(D) \setminus C$ into sets $Z_0, \ldots, Z_t$, which we call *zones*, as follows. For every $v \in V(D) \setminus C$, let i be the largest index such that v is seen by at least one vertex in B_i. Then vertex v is assigned to zone Z_i. Otherwise, we assign v to zone Z_0. This partition is unique and can be done in time $poly(n)$. As in the case of the bags and zones used in Section 7.3, these bags and zones exhibit useful properties. The proofs we present here are similar, but much simpler.

CLAIM 8.11

> *Let $C = \{B_1, \ldots, B_t\}$ be a chain of bags, and let $Z_0, Z_1, \ldots, Z_t$ be a partition of the vertices in $V(D) \setminus C$. For every i, the following properties hold:*
>
> *(a) $B_i \to B_{i+r}$ for every $r \geqslant 1$,*
>
> *(b) $Z_i \to B_{i+r}$ for every $r \geqslant 1$,*
>
> *(c) $B_i \to Z_{i+r}$ for every $r \geqslant 2$,*
>
> *(d) $Z_i \to Z_{i+r}$ for every $r \geqslant 3$.*

Proof. Property (a) holds for $r = 1$ by definition of a chain of bags. Suppose that (a) holds for $r - 1 > 1$, and suppose that there is an arc uv with $u \in B_{i+r}$ and $v \in B_i$. Since B_{i+1} is a bag, there is $x \in B_{i+1}$ such that x is a common neighbor of u and v. Then by induction hypothesis, vx, xu are arcs, and so vxu is a copy of C_3, a contradiction. Hence (a) holds for r.

Property (b) holds for all $r \geqslant 1$ by the partitioning criterion of vertices into zones.

To prove property (c), suppose that there is an arc zv with $z \in Z_{i+r}$ and $v \in B_i$ for some $r \geqslant 2$. Then there is $u \in B_{i+r}$ such that uz is an arc by the partitioning criterion of vertices into zones. Since B_{i+1} is a bag, there is $x \in B_{i+1}$ such that x is a common neighbor of u, v, z. By property (a), vx and xu are arcs. If xz is an arc, then vxz is a copy of C_3. Otherwise, zx is an arc, and so xuz is a copy of C_3. Either way, we reach the contradiction, and so (c) holds for every $r \geqslant 2$.

To prove property (d), suppose that there is an arc uv with $u \in Z_{i+r}$ and $v \in Z_i$ for some $r \geqslant 3$. Since B_{i+1} is a bag, there is $x \in B_{i+1}$ such that x is a common neighbor of all u and v. By property (b), vx is an arc, and by property (c), xu is an arc. Hence vxu is a copy of C_3, a contradiction. Hence (d) holds for every $r \geqslant 3$. □

We now show how to find a chain of poor bags, which can be colored using Claim 8.9.

FIND-CHAIN(D, B)

(i) If there is $v \in B$ such that both $N^+(v) \cap B$ and $N^-(v) \cap B$ are bags of D, then:

Return $\big($FIND-CHAIN$(D, N^-(v) \cap B), v,$ FIND-CHAIN$(D, N^+(v) \cap B)\ \big)$.

(ii) Otherwise, return B.

The routine FIND-CHAIN(D, B) returns $B_1, v_1, B_2, \ldots, v_{s-1}, B_t$, i.e. a sequence of sets B_i alternating with vertices v_i. We say that the sequence $B_1, B_2, \ldots, B_t$ is the chain of poor bags output by the procedure FIND-CHAIN(D, B). Later on, we will use the vertices v_i in the output sequence to facilitate the coloring of vertices outside the chain. Observe that if B is a poor bag or if B is not a bag, then FIND-CHAIN(D, B) returns a single set, namely B.

CLAIM 8.12

If bag $B \subseteq V(D)$ is not poor, then FIND-CHAIN(D, B) *returns a chain of poor bags $B_1, \ldots, B_t$ for some $t \geqslant 2$ in time $poly(n)$.*

Proof. Let $B_1, \ldots, B_t$ be the chain of poor bags output by FIND-CHAIN(D, B). The bags in this chain are pairwise disjoint. From Step 1 of FIND-CHAIN, it follows that each B_i is a bag. Furthermore, each B_i must be poor; otherwise, Step 1 would be applied to B_i to return poor bags inside B_i. Observe that for every pair of consecutive bags B_i, B_{i+1} in the chain, $B_{i+1} \subseteq N^+(v_i)$ and $B_i \subseteq N^-(v_i)$. If there is an arc xy with $x \in B_{i+1}$ and $y \in B_i$, then $v_i xy$ is a copy of C_3, a contradiction. Hence $B_1, \ldots, B_t$ is a chain of poor bags by definition. The procedure FIND-CHAIN(D, B) runs in time $poly(n)$. □

CLAIM 8.13

If FIND-CHAIN(D, B) *returns a chain of t poor bags, then we can color B with $8\alpha\beta + (t-1) \cdot \beta$ colors in time $poly(n) + time(\alpha - 1, |B|)$.*

Proof. Suppose that FIND-CHAIN(D, B) returns $B_1, v_1, B_2, \ldots, v_{t-1}, B_t$ and that $C = \{B_1, \ldots, B_t\}$ is the chain of poor bags output by the procedure. Since each B_i is a poor bag, it can be colored using Claim 8.9. By Claim 8.11, $B_i \to B_j$ for every $i < j$, and so we can color the vertices in C using $8\alpha \cdot \beta$ colors in time $poly(n) + time(\alpha - 1, |C|)$ time.

Observe that each $v \in B \setminus C$ is either (i) some v_i in the output sequence returned by FIND-CHAIN(D, B), or (ii) belongs to $N^o(v_i)$ for some v_i in this output sequence. For any $v \in V(D)$, the set $N^o(v)$ is $(\alpha - 1)$-dense. Therefore, the vertices in $\bigcup_{i=1}^{t-1} N^o(v_i)$ can be colored with $(t-1)\beta$ colors. Note that v_i can be colored with an arbitrary color from the color palette used for $N^o(v_i)$. This coloring can be found in time $poly(n) + time(\alpha - 1, |B \setminus C|)$. □

COROLLARY 8.14

Either FIND-CHAIN(D, B) *returns a chain of t poor bags, or B can be colored using $8\alpha\beta + (t-2) \cdot \beta$ colors.*

We now have the tools to outline our main coloring algorithm.

COLOR-DIGRAPH(D)

If D is acyclic, color D with one color and terminate.

Otherwise:

(i) Run FIND-CHAIN$(D, V(D))$ and let $C := \{B_1, \ldots, B_t\}$ denote the chain of poor bags that it outputs.

(ii) Assign each vertex in $V(D) \setminus C$ to a zone Z_i for $i \in [0, t]$.

(iii) While FIND-CHAIN(D, Z_i) returns a chain of poor bags $B'_1, B'_2, \ldots, B'_k$ for $k \geqslant 3$:

  (i) Update chain: $C := \{B_1, \ldots, B_{i-2}, B'_1, B'_2, \ldots, B'_k, B_{i+1}, \ldots, B_t\}$.

  (ii) Re-assign each vertex in $V(D) \setminus C$ to a zone.

(iv) Color all vertices in the chain C with $8\alpha\beta$ colors.

(v) Color all vertices in the zones of C with $3(8\alpha + 1)\beta$ colors.

CLAIM 8.15

COLOR-DIGRAPH(D) *colors D with at most $35\alpha \cdot \beta$ colors.*

Proof. If $V(D)$ is a poor bag, then we can apply Claim 8.9. Hence we may suppose that $V(D)$ is not a poor bag.

Note that the updated chain C resulting from Step 3 (a) is still a chain of poor bags, due to properties (b) and (c) from Claim 8.11. After Step 3 finishes, the chain C is *maximal* in that the procedure FIND-CHAIN(D, Z_i) will not find a chain of three poor bags in any zone Z_i. Using Corollary 8.14, we can therefore color each zone

using $(8\alpha+1)\beta$ colors. Applying property (d) from Claim 8.11, we can use at most $(24\alpha+3)\cdot\beta$ colors to color all vertices in $V(D)\setminus C$. Each bag B_i in the chain C is a poor bag, so we can color B_i with $8\alpha\beta$ colors by Claim 8.9. Moreover, since $B_i \to B_j$ for every $i<j$ (by property (a) from Claim 8.11), we need $8\alpha\beta$ colors to color the entire chain C. Thus, we can color D with $(32\alpha+3)\beta \leqslant 35\alpha\beta$ colors. □

CLAIM 8.16

The procedure COLOR-DIGRAPH(D) *uses* $35^{\alpha-1}\alpha!$ *colors.*

Proof. We proceed by induction on α. If $\alpha = 1$, then we use one color, since every C_3-free tournament is acyclic. Suppose that the algorithms colors each $(\alpha-1)$-dense C_3-free digraph D' using at most β colors. Then by Claim 8.15, the procedure COLOR-DIGRAPH(D) colors D with at most

$$35\alpha\beta \;=\; 35\alpha\cdot 35^{\alpha-2}(\alpha-1)! \;=\; 35^{\alpha-1}\alpha!$$

colors. □

CLAIM 8.17

COLOR-DIGRAPH(D) *runs in time* $poly(n)$.

Proof. We now analyze the running time. Finding a maximal chain C takes $poly(n)$ time, as does the procedure of partitioning the vertices in $V(D)\setminus C$ into zones. Once we have found this partitioning, we can color the vertices in C in time $poly(n) + time(\alpha-1,|C|)$ using Claim 8.9, and we can color each zone Z_i in time $poly(n)+(\alpha-1,|Z_i|)$ using Claim 8.13 and Corollary 8.14. So applying Claim 8.6, the total running time of COLOR-DIGRAPH(D) is at most $poly(n)+time(\alpha-1,n)$. This leads to the following recurrence relation:

$$\begin{aligned} time(\alpha,n) &= poly(n)+time(\alpha-1,n)\\ &= \alpha\cdot poly(n). \end{aligned}$$

Since $\alpha \leqslant n$, we have that $time(\alpha,n) = poly(n)$.

We note that our algorithm is actually just partitioning $V(D)$ into disjoint $(\alpha-1)$-dense subsets. In other words, it first partitions the vertices in $V(D)$ into sets where each set is a poor bag or not a bag (for example, a zone Z_i is either a poor bag or not a bag). Then, it further partitions these sets into $(\alpha-1)$-dense sets. (For example, by Claim 8.7, a set that is not a bag can be partitioned into three disjoint sets, each is $(\alpha-1)$-dense. Similarly, in the proof of Claim 8.9, a poor bag is partitioned into 8α disjoint sets, each is $(\alpha-1)$-dense.) Once the subgraphs on these induced subsets are colored (recursively) using β colors, then certain subsets are allowed to use the same color palette, and these color palettes can be coordinated in time $poly(n)$. The initial partitioning procedure and the final coordinating procedure require $poly(n)$ time, while the recursive coloring requires $time(\alpha-1,n)$ time. □

Theorem 8.2 follows from Claims 8.15, 8.16 and 8.17.

Finally, we remark that Theorem 8.2 yields a bound on the size of a maximum acyclic subgraph of α-dense C_3-free digraphs.

THEOREM 8.18

Let $D = (V, A)$ be an α-dense C_3-free digraph. Then D contains an acyclic subset of arcs, $A' \subseteq A$, with cardinality $|A'| \geqslant |A| \cdot (\frac{1}{2} + c_\alpha)$, where c_α is a constant depending on α.

Proof. Since D can be colored with $35^{\alpha-1}\alpha!$ colors, there is a color class contains at least $\frac{n}{35^{\alpha-1}\alpha!}$ vertices, and by Turan's Theorem, its induced subgraph contains at least

$$\frac{n}{2 \cdot 35^{\alpha-1}\alpha!}\left(\frac{n}{\alpha \cdot 35^{\alpha-1}\alpha!} - 1\right)$$

arcs. Thus, D contains a maximum acyclic subgraph of size at least

$$\frac{|A|}{2} + \frac{n}{4 \cdot 35^{\alpha-1}\alpha!}\left(\frac{n}{\alpha \cdot 35^{\alpha-1}\alpha!} - 1\right).$$

Let c be an absolute constant. Then

$$c_\alpha = \frac{1}{c \cdot \alpha \cdot (35^{\alpha-1}\alpha!)^2}$$

satisfies the theorem. □

8.2 Domination in triangle-free digraphs

Before concluding this thesis, we will discuss briefly about dominating sets in dense C_3-free digraphs. As we shown through previous chapters, there is a tight bond between domination and coloring in digraphs. Another example is that *for any α-dense digraph D, we have $\gamma(D) \leqslant \alpha\chi(D)$*, which follows from the fact that in a proper coloring, each color class forms an induced acyclic digraph and every acyclic digraph has a stable dominating set (by Proposition 6.16). Thus, combining with Theorem 8.2, we have that α-dense C_3-free digraph D has the domination number at most $\alpha 35^{\alpha-1}\alpha!$. However, a better bound can be achieved, where $35^{\alpha-1}$ is dropped.

THEOREM 8.19

For every positive integer α and any α-dense C_3-free digraph D, we have $\gamma(D) \leqslant \alpha\alpha!$.

For small cases, we can provide better bounds, for example, the tight bound $\gamma(D) \leqslant 3$ when $\alpha = 2$, or $\gamma(D) \leqslant 8$ when $\alpha = 3$. Similar to Conjecture 8.1, we also believe that a polynomial bound for the size of dominating sets in dense C_3-free digraphs.

CONJECTURE 8.20

There is an integer ℓ such that for every positive integer α, if D is an α-dense C_3-free digraph, then D has a dominating set of size α^ℓ.

We now prove Theorem 8.19.

Proof of Theorem 8.19. Let $h(1) = 1$ and $h(k) = k \cdot (h(k-1)+1)$ for every $k > 1$. We proceed by induction on α that $\gamma(D) \leqslant h(\alpha)$. If $\alpha = 1$, then D is a C_3-free tournament, and so D is acyclic. Then there is a vertex that dominates $V(D)$. So $h(1) = 1$.

If $\alpha > 1$, we first apply Proposition 8.3 to find an acyclic dominating set A of D. Then we apply Proposition 6.16 to find a dominating set $S \subseteq A$ of A. Note that $|S| \leqslant \alpha$ since S is stable. Let Y denote the set of vertices in $V(D) \setminus A$ that are not dominated by any vertex in S.

We observe that each vertex $y \in Y$ belongs to $N^o(s)$ for some $s \in S$. Suppose this were not the case. Then a vertex $y \in Y$ has arcs to all $s \in S$. Since A dominates V, there is some $a \in A \setminus S$ such that $ay \in E(D)$. Since S dominates A, there is $s \in S$ such that $sa \in E(D)$. This yields the triangle ysa, which is a contradiction to D being C_3-free. Hence, for every $y \in Y$, there is $s \in S$ such that $ys \notin E(D)$. By the definition of Y, it follows that $sy \notin E(D)$. We conclude that s and y are independent and therefore $y \in N^o(s)$.

We partition Y into $|S|$ subsets, where for each $s \in S$, Y_s is a subset of $N^o(s)$. Note that $D[N^o(s)]$ is $(\alpha-1)$-dense, then so is $D[Y_s]$. By the induction hypothesis, there is a set B_s of size at most $h(\alpha-1)$ dominating Y_s. Let $B = S \cup \left(\bigcup_{s\in S} B_s\right)$. Then $|B| \leqslant |S| + |S| \cdot h(\alpha-1) \leqslant h(\alpha) < \alpha\alpha!$.

Observe that $Y = \bigcup_{s\in S} Y_s$ is dominated by $\bigcup_{s\in S} B_s$, and $V \setminus Y$ is dominated by S by definition of Y. Thus, B is a dominating set of D, which proves the theorem. □

Conclusion

Throughout the thesis, we have studied the relation between patterns and large graphs. Let us sum up the results and propose some directions for further work.

In Part I, we discussed about the sufficient conditions so that a large graph can be decomposed into copies of a given pattern. In Chapter 2, we proved the Bárat-Thomassen conjecture that every highly edge-connected graph can be decomposed into copies of a given tree. We provided a large exponential upper bound on the edge-connectivity which holds for all patterns. One direction for future work could be improving this general bound, or bounds for specific patterns. Another idea could be finding some lower bounds on the edge-connectivity for specific patterns. For example, we showed in Chapter 3 that when the pattern is P_9, the bound on the edge-connectivity must be at least 3. These lower bounds could give insight to the true sharp bound of the Barát-Thomassen conjecture.

One could also investigate a counterpart of the Barát-Thomassen conjecture for patterns which are not a tree (i.e. containing a cycle). There are numerous works on that topic, mainly focusing on the cases where the pattern is a cycle or a complete graph, for example, [Wil75, BKLO15, BC15]. However, it is easy to notice that in this case, high edge-connectivity is not sufficient; the large graph needs to have small *girth*, too. Indeed, suppose that the pattern contains a cycle of length ℓ, then the large graph must contains many cycles of length ℓ as well.

In Chapter 3, we showed that when the pattern is a path, it is possible to replace the edge-connectivity condition in the Barát-Thomassen conjecture by a cheaper one – the minimum degree. Whether it is possible to improve the Barát-Thomassen conjecture when the pattern is a tree is still open.

In Chapter 4, we proved a conjecture by Häggkvist and Kriesell that every Eulerian graph with high minimum degree admits an ℓ-step self-avoiding Eulerian tour. Which immediately implies that every Eulerian graph with high minimum degree can be decomposed into copies of a given path, i.e. replacing the high edge-connectivity condition of the Barát-Thomassen conjecture by Eulerianity and high minimum degree. Oksimets [Oks97] showed that 6 is a sharp bound for $\ell = 3$. Finding sharp

bounds for other patterns could be a direction for future work.

In Chapter 5, we proved Jaeger-Linial-Payan-Tarsi's conjecture on additive bases in the case the vectors has support at most 2. This results implies several results on flows and orientations in graphs. Improving the bound on the number of bases will subsequently improve the bound on the edge-connectivity condition of its corollaries on flows and orientations in graphs. Since a vector of support 2 can be mapped into an edge, a result similar where vectors have support at most 3 could imply similar corollaries for hypergraphs with edge size at most 3. A general result without bounded support size would give results on flows and orientations in general hypergraphs.

Part II was dedicated to the tournament version of the Erdős-Hajnal conjecture. In Chapter 6, we proved a conjecture of Berger et al that if a tournament is nowhere locally colorful, then it is not globally colorful. It is natural to ask the reverse question: If a large tournament is colorful, is it true that it is locally colorful somewhere, i.e. can we say that it must contain a small but colorful subtournament? This turns out to be false; an counterexample could be found at [HLTW17]. In [HLTW17], we indeed presented a stronger result that the size of local domination sets dictates the global colorfulness of a tournament. This again shows the close tie between domination and coloring in graphs. Hence, it is natural to ask whether the size of local domination sets dictates the size of the global domination set. This question is still open.

In Chapter 7, we discussed the general Erdős-Hajnal conjecture on digraphs and extended the result of Berger et al [BCC$^+$13] to dense digraphs, and discussed in detail the case where the pattern is the directed-triangle in Chapter 8. An imminent direction for future work could be proving the general Erdős-Hajnal conjecture for some small patterns, such as the directed-triangle. Another idea could be replacing the dense condition by some weaker ones, such as high average degree or high minimum degree.

Contributions

The results presented in this thesis are from:[1]

[1] Julien Bensmail, Ararat Harutyunyan, Tien-Nam Le, Martin Merker, and Stéphan Thomassé. A proof of the barát–thomassen conjecture. *Journal of Combinatorial Theory, Series B*, 124:39–55, 2017.

[2] Julien Bensmail, Ararat Harutyunyan, Tien-Nam Le, and Stéphan Thomassé. Edge-partitioning a graph into paths: beyond the barát- thomassen conjecture. *Combinatorica*, pages 1–25, 2018.

[3] Tien-Nam Le. Locally self-avoiding Eulerian tours. *Journal of Combinatorial Theory, Series B*, to appear, and *LAGOS* 2017.

[4] Louis Esperet, Rémi de Joannis de Verclos, Tien-Nam Le, and Stéphan Thomassé. Additive bases and flows in graphs. *SIAM Journal of Discrete Mathematics*, 32:534–542, 2018, and *Eurocomb* 2017.

[5] Ararat Harutyunyan, Tien-Nam Le, Stéphan Thomassé, and Hehui Wu. Coloring tournaments: from local to global. *arXiv preprint* arXiv:1702.01607, 2017.

[6] Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé. Coloring dense digraphs. *Combinatorica*, to appear, and *Eurocomb* 2017.

[7] Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé. Domination and fractional domination in digraphs. *Electronic Journal of Combinatorics*, to appear.

[1] All these papers (except [3]) have not been accepted in whole, or in part, for any other degree or diploma of the author. A part of [3] appeared in his master report.

Bibliography

[Ade95] Tobias Adelgren. Triangle-free eulerian tours in graphs with maximum degree at most 4. *Discrete mathematics*, 138(1-3):5–14, 1995.

[AHKT18] Noga Alon, Dan Hefetz, Michael Krivelevich, and Mykhaylo Tyomkyn. Edge-statistics on large graphs. *arXiv preprint arXiv:1805.06848*, 2018.

[ALM91] Noga Alon, Nathan Linial, and Roy Meshulam. Additive bases of vector spaces over prime fields. *J. Comb. Theory, Ser. A*, 57(2):203–210, 1991.

[ANR96] Noga Alon, Melvyn B Nathanson, and Imre Ruzsa. The polynomial method and restricted sums of congruence classes. *Journal of Number Theory*, 56(2):404–417, 1996.

[APS01] Noga Alon, János Pach, and József Solymosi. Ramsey-type theorems with forbidden subgraphs. *Combinatorica*, 21(2):155–170, 2001.

[BC15] Darryn Bryant and Nicholas J Cavenagh. Decomposing graphs of high minimum degree into 4-cycles. *Journal of Graph Theory*, 79(3):167–177, 2015.

[BCC+13] Eli Berger, Krzysztof Choromanski, Maria Chudnovsky, Jacob Fox, Martin Loebl, Alex Scott, Paul Seymour, and Stephan Thomassé. Tournaments and colouring. *Journal of Combinatorial Theory, Series B*, 103(1):1–20, 2013.

[BG14] János Barát and Dániel Gerbner. Edge-decomposition of graphs into copies of a tree with four edges. *The Electronic Journal of Combinatorics*, 21(1):P1–55, 2014.

[BKLO15] Ben Barber, Daniela Kühn, Allan Lo, and Deryk Osthus. Edge-decompositions of graphs with high minimum degree. *Electronic Notes in Discrete Mathematics*, 49:115–121, 2015.

[BM+76] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Citeseer, 1976.

[BMOW16] Fábio Botler, Guilherme O Mota, Marcio TI Oshiro, and Yoshiko Wakabayashi. Decomposing highly connected graphs into paths of length five. *Discrete Applied Mathematics*, 2016.

[BMOW17] Fábio Botler, Guilherme O Mota, Marcio TI Oshiro, and Yoshiko Wakabayashi. Decomposing highly edge-connected graphs into paths of any given length. *Journal of Combinatorial Theory, Series B*, 122:508–542, 2017.

[BT06] János Barát and Carsten Thomassen. Claw-decompositions and tutte-orientations. *Journal of Graph Theory*, 52(2):135–146, 2006.

[CCS14] Krzysztof Choromanski, Maria Chudnovsky, and Paul Seymour. Tournaments with near-linear transitive subsets. *Journal of Combinatorial Theory, Series B*, 109:228–249, 2014.

[Chu14] Maria Chudnovsky. The erdös–hajnal conjecture—a survey. *Journal of Graph Theory*, 75(2):178–190, 2014.

[CKL+18] Maria Chudnovsky, Ringi Kim, Chun-Hung Liu, Paul Seymour, and Stéphan Thomassé. Domination in tournaments. *Journal of Combinatorial Theory, Series B*, 130:98–113, 2018.

[CXWW14] Wanyun Cui, Yanghua Xiao, Haixun Wang, and Wei Wang. Local search of communities in large graphs. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 991–1002. ACM, 2014.

[Dav35] Harold Davenport. On the addition of residue classes. *Journal of the London Mathematical Society*, 1(1):30–32, 1935.

[Die05] Reinhard Diestel. *Graph theory (Graduate texts in mathematics)*, volume 173. Springer Heidelberg, 2005.

[Edm73] Jack Edmonds. Edge-disjoint branchings. *Combinatorial algorithms*, 1973.

[EH89] Paul Erdös and András Hajnal. Ramsey-type theorems. *Discrete Applied Mathematics*, 25(1-2):37–52, 1989.

[Erd47] Paul Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[Fau07] Katherine Faust. 7. very local structure in social networks. *Sociological Methodology*, 37(1):209–256, 2007.

[Hag89] Rolald Haggkvist. Decompositions of regular bipartite graphs. In *Surveys in combinatorics, 1989 (Siemons)*, volume 141, pages 115–147. Cambridge Univ. Press Cambridge, 1989.

[HJJ+18] Yanqing Hu, Shenggong Ji, Yuliang Jin, Ling Feng, H Eugene Stanley, and Shlomo Havlin. Local structure can identify and quantify influential global spreaders in large scale social networks. *Proceedings of the National Academy of Sciences*, page 201710547, 2018.

[HLTW17] Ararat Harutyunyan, Tien-Nam Le, Stéphan Thomassé, and Hehui Wu. Coloring tournaments: from local to global. *arXiv preprint arXiv:1702.01607*, 2017.

[HM12] Ararat Harutyunyan and Colin McDiarmid. Large acyclic sets in oriented graphs. *Unpublished note*, 2012.

[HYL17] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

[Jac93] Bill Jackson. On circuit covers, circuit decompositions and euler tours of graphs. In *Surveys in combinatorics, 1993 (Keele)*, volume 187, pages 191–210. Cambridge Univ. Press Cambridge, 1993.

[Jae79] François Jaeger. Flows and generalized coloring theorems in graphs. *Journal of Combinatorial Theory, Series B*, 26(2):205–216, 1979.

[JLM17] Hua Jiang, Chu-Min Li, and Felip Manya. An exact algorithm for the maximum weight clique problem in large graphs. In *AAAI*, pages 830–838, 2017.

[JLPT92] François Jaeger, Nathan Linial, Charles Payan, and Michael Tarsi. Group connectivity of graphs—a nonhomogeneous analogue of nowhere-zero flow properties. *Journal of Combinatorial Theory, Series B*, 56(2):165–182, 1992.

[Kri11] Matthias Kriesell. *Fourth international conference on combinatorics, graph theory, and applications*, 2011.

[KT17] Tereza Klimošová and Stéphan Thomassé. Decomposing graphs into paths and trees. *Electronic Notes in Discrete Mathematics*, 61:751–757, 2017.

[LL17] Hong-Jian Lai and P. Li. Additive bases and strongly z_2s+1-connectedness. *Manuscript*, 2017.

[LLDM09] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.

[LTWZ13] László Miklós Lovász, Carsten Thomassen, Yezhou Wu, and Cun-Quan Zhang. Nowhere-zero 3-flows and modulo k-orientations. *Journal of Combinatorial Theory, Series B*, 103(5):587–598, 2013.

[McD02] Colin McDiarmid. Concentration for independent permutations. *Combinatorics, Probability and Computing*, 11(2):163–178, 2002.

[Mer17] Martin Merker. Decomposing highly edge-connected graphs into homomorphic copies of a fixed tree. *Journal of Combinatorial Theory, Series B*, 122:91–108, 2017.

[MR13] Michael Molloy and Bruce Reed. *Graph colouring and the probabilistic method*, volume 23. Springer Science & Business Media, 2013.

[NP12] T Niinimki and Pekka Parviainen. Local structure discovery in bayesian networks. In *28th Conference on Uncertainty in Artificial Intelligence, UAI 2012; Catalina Island, CA; United States; 15 August 2012 through 17 August 2012*, pages 634–643, 2012.

[NW60] C St JA Nash-Williams. On orientations, connectivity and odd vertex pairings in finite graphs. *Canad. J. Math*, 12(555-567):8, 1960.

[Oks97] Natalia Oksimets. *Triangle-free eulerian tours in graphs with minimum degree at least 6*. Tekniska Högskolan i Luleaa. Department of Mathematics, 1997.

[Sey81] Paul D Seymour. Nowhere-zero 6-flows. *Journal of Combinatorial Theory, Series B*, 30(2):130–135, 1981.

[SSTF12] Michael Stiebitz, Diego Scheide, Bjarne Toft, and Lene M Favrholdt. *Graph Edge Coloring: Vizing's Theorem and Goldberg's Conjecture*, volume 75. John Wiley & Sons, 2012.

[Ste59] Richard Stearns. The voting problem. *The American Mathematical Monthly*, 66(9):761–763, 1959.

[Ste76] R Steinberg. Grötzsch's theorem dualized, m. *Math Thesis, University of Waterloo, Ontario, Canada*, 1976.

[Tho08a] Carsten Thomassen. Decompositions of highly connected graphs into paths of length 3. *Journal of Graph Theory*, 58(4):286–292, 2008.

[Tho08b] Carsten Thomassen. Edge-decompositions of highly connected graphs into paths. In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, volume 78, page 17. Springer, 2008.

[Tho12] Carsten Thomassen. The weak 3-flow conjecture and the weak circular flow conjecture. *Journal of Combinatorial Theory, Series B*, 102(2):521–529, 2012.

[Tho13a] Carsten Thomassen. Decomposing a graph into bistars. *Journal of Combinatorial Theory, Series B*, 103(4):504–508, 2013.

[Tho13b] Carsten Thomassen. Decomposing graphs into paths of fixed length. *Combinatorica*, 33(1):97–123, 2013.

[Tho14] Carsten Thomassen. Graph factors modulo k. *Journal of Combinatorial Theory, Series B*, 106:174–177, 2014.

[Tut54] William Thomas Tutte. A contribution to the theory of chromatic polynomials. *Canad. J. Math*, 6(80-91):3–4, 1954.

[Wil75] Richard M Wilson. Decompositions of complete graphs into subgraphs isomorphic to a given graph. In *Proc. 5th British Combinatorial Conf., 1975*, 1975.